DOCUMENT RESUME

ED 319 761 TM 014 942

AUTHOR Mildner, Telma
TITLE A Statewide Assessment and Evaluation of Fourth Grade Mathematics Delivery System.
INSTITUTION Ministry of Education and Culture, Jerusalem (Israel). Israel Curriculum Center.
PUB DATE 89
NOTE 254p.
PUB TYPE Books (010) -- Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC11 Plus Postage.
DESCRIPTORS Accountability; Achievement Tests; *Curriculum Evaluation; Educational Assessment; Educational Improvement; *Elementary School Mathematics; Feedback; Foreign Countries; *Grade 4; Intermediate Grades; Mathematics Curriculum; *Mathematics Education; *National Surveys; Program Evaluation; Student Attitudes; Teacher Attitudes; Teacher Evaluation; *Teaching Conditions
IDENTIFIERS *Israel

ABSTRACT

The Israel Ministry of Education and Culture (Jerusalem) has attempted to establish a feedback system to provide information about the functioning of the educational system in Israel. Feasibility of such a system and ways to ensure its effectiveness were studied through an assessment of the teaching of mathematics in grade 4. Data were collected about teaching conditions and student achievement in a stratified probability sample of 99 schools. Teachers provided information about teaching conditions, and 2,843 students answered a questionnaire about studying mathematics and completed a 40-item achievement test. Individual and class statistics were derived and achievement profiles were portrayed. Teachers were given the information in an effort to improve their teaching. Data were also analyzed to describe three aspects of fourth-grade mathematics education: (1) intended curriculum, through analysis of the National Syllabus and widely used curricula; (2) implemented curriculum, through examination of salient components; and (3) attained curriculum, through assessment of achievement data. The methodology may provide an illustrative example for other countries interested in accountability. A 288-item list of references is included. Three appendices containing information on the tests and the study instruments in Hebrew are provided. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319761

A STATEWIDE ASSESSMENT AND EVALUATION OF FOURTH GRADE MATHEMATICS DELIVERY SYSTEM

סקר הישגים ועמדות, מתימטיקה כיתה ד'

TELMA MILDNER

Jerusalem
Ministry of Eduction and Culture
Pedagogical Administration
Israel Curriculum Center
1989

משרד החינוך והתרבות
המינהל הפדגוגי
האגף לתכניות לימודים
ירושלים התשמ"ט

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TELMA MILDNER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

TM018482

A STATEWIDE ASSESSMENT AND EVALUATION OF FOURTH GRADE MATHEMATICS DELIVERY SYSTEM

By
TELMA (BENDER) MILDNER

1988

JERUSALEM, MINISTRY OF EDUCATION AND CULTURE
ISRAEL CURRICULUM CENTER

Table of Contents

	Page
Abstract	VIII
List of Tables	XI
List of Figures and Graphs	XIII
List of Exhibits	XIV
Chapter One	
The Theoretical Framework	1
The Background of the Study	1
Evaluation and assessment	4
Changing Conceptions About Examining Educational Achievements	4
Trends in Assessment	7
Trends in Evaluation Studies	12
Links Between Evaluation and Assessment	18
Trends in Elementary School Mathematics Curricula	19
The Concept of Minimum Competency	30
Approaches to Defining Minimum Competencies	31
The Emergence of Minimum Competency Programs in the 1970s	32
Criticism of the Minimum Competency Testing	35
Approaches to Minimum Competency in Israel	36
Chapter Two	
The Operational Framework	40
The Goals of the Study	40
Substantive Aspects	40
Methodological Aspects	42
The Management of the Study	42

Page

The Population and the Sample 43
The Population 43
The Sample 44
The Stratifying Variables 44

The Data Collection Instruments 50
Teacher's Questionnaire 50
Pupil's Questionnaire 51
The Mathematics Achievement Test 51
The Judgmental-Empirical Approach to Modifying the Minimum Competency Requirements 57

Data Collection 60

Data Analysis 62

Chapter Three
Mathematics Intended Curricula 64

The National Syllabus 64

Widely-Used Instructional Programs 66
The "New Mathematics" Instructional Program 68
The "One Two and... Three" Instructional Program 70
Simple Math: The Television Fourth-Grade Mathematics Program 74

Chapter Four
The Implemented Curriculum 78

The Background of Learners 78

The Teacher 83
The Professional Educational and Experience of the Teacher 83

Page

The Autonomy of the Teacher 87
Attitude Toward the Program Used in the Class 90

Time-Related Variables 94

Implementation of the Program at Classroom Level 97

Opportunity to Learn Mathematics Topics 106
Differences Between Schools 109
Differences Between Topics 109
Differences Between the Level of Coverage of Various Items 112
Items Covered in the Whole System 113

Chapter Five
Attainment of Intended Instructional Goals 115

Overall Results: Individual Scores 116

Overall Results: School Mean Scores 120

The Distribution of Low Achieving Students Across Classes 123

Results Related to the Subsections of the Test 124
The Prescribed Grade Level of Topics 125
Word Problems Versus Numerical Exercises 126
Specific Topics 127

Results Related to the Individual Items 128
The Test Items Known by Almost All Pupils 129
The Test Items Known by Few Pupils Only 130
Discrimination Among Schools Versus Discrimination Among Individuals 130

Page

How Much Mathematics do Fourth Grade Pupils Know 133

Achievement Profiles 136
Achievement Profile of the Whole Population 138
Profiles of Users of Various Textbooks 139
School Profiles 143
Individual Profiles 145

Attitude Toward Mathematics 149

Chapter Six
Factors Associated with Outcomes of Mathematics Teaching 152

Achievement in the Six Administrative School Districts 152

Gender and Achievement in Mathematics 153
Studies About the Relationship Between Gender and Mathematics Achievement 153
Sex Related Findings in the Present Study ... 155

Schools with Different Social Priority Rating ... 159

Users of Various Programs 161

Chapter Seven
The Utilization of Evaluation Results 163

Features of the Study which Facilitated Utilization 167
Planning for Utilization 167
Orientation Toward the User 168
Users' Personal Involvement 168
Multiple Clientele 169

Page

Expected and Actual Decisions Taken
Based on the Study Results 170
Utilization by Classroom Teachers 171
Utilization by Regional School Supervisors .. 172
Utilization by Curriculum Developers 174
Other Audiences 175

Increasing the Utilization of the Results 176

Chapter Eight
Summary and Recommendations..... 179

Summary 179
A Multipurpose Study Responding to
Multiple Clientele 180
The Intended Curriculum 181
The Implemented Curriculum 182
The Attained Curriculum 185
The Utilization Aspect 187

Recommendations 189
Recommendations to Producers of
Instructional Programs 189
Recommendations to Planners of
Teacher Education Programs 190
Recommendations to the National and
Regional Educational Authorities 190
Recommendations to Superintendents
and Supervisors 191
Recommendations to Principals
and Teachers 191
Recommendations to Evaluation
Professionals 193

Bibliography 194

Appendices 221

Abstract

Over the last ten years attempts have been made by the Ministry of Education and Culture to establish a Feedback System, which will provide systematic information about various aspects of the functioning of the educational system in Israel. A committee established by the Chief Scientist of the Ministry has issued a series of recommendations about desired parameters of the system, and suggested that high priority should be given to the systematic provision of information about achievements in various subject areas and at different grade levels of the school. The present study is an attempt to explore the feasibility of establishing and maintaining a multipurpose Feedback System, and to find out in what ways its effectiveness can be ensured. Such a Feedback System will serve a multiple clientele and will fulfill a dual role in the educational system: evaluating the variety of programs and teaching strategies within the framework of teaching a given subject, and assessing the educational outcomes at individual and group (aggregate) levels, with the aim of providing diagnostic information. At aggregate level, care will be taken to supply information needed by municipalities, administrative school districts, supervisors, schools principals, teachers and other interested parties.

In analysing experiences accumulated around the world in the operation of Feedback Systems in the field of education, an attempt was made to establish a conceptual framework and to recommend an operational strategy, which may eliminate pitfalls observed in other countries.

The findings of the study may guide educational authorities into increasing the utility of the planned Feedback System for various user groups, such as teachers, school administrators, curriculum developers, educational leaderships and other interested parties. The study focused on exploring the unique information needs of the educational system in Israel, but at the same time attention was paid to defining universal characteristics of the Feedback System in a way which may be applicable to such systems in general. The author, being a faculty member of a University in Brazil, intends to present this study as a model for establishing a Feedback System in Brazilian States and as starting point for cross-national cooperation between educational systems which maintain Feedback Systems.

The theoretical contribution of the study touches on demonstrating the interdependence between what is traditionally referred to as assessment study and what is labelled as program evaluation study. The study provided evidence that traditional conceptions about differences between assessment and evaluation studies are not necessarily valid, and that there are ways of reducing an erroneously conceived dichotomy between these two study archetypes. This means that one can derive information from a single set of data to fulfill roles which, in the past, have usually been distinctively attributed either to assessment or to evaluation studies.

To demonstrate the interdependence between assessment and progam evaluation, a large scale national study was carried out to examine the state of the art of teaching of mathematics in the fourth grade of the Israeli elementary school.

Data was collected in regard to conditions of teaching mathematics and pupils' attainments in fourth grade mathematics in a stratified probability sample of 99 schools. The fourth grade mathematics teachers provided information about the conditions of teaching fourth grade mathematics in their schools, and 2843 pupis answered both a questionnaire dealing with feelings and experierces related to studying mathematics and a 40-item mathematics achievement test.

The data collected was analysed in a way which served given information needs of school teachers related to pupils' achievement levels: individual and class statistics were derived, achievement profiles were portrayed, and teachers were guided to use them for improving their teaching.

Teachers used the information as a basis for validating their overall scores assigned to the pupils, to facilitate the grouping of pupils within the class, and only to a very limited extent for assigning differential remedial treatment to the students.

The data was also analysed in a way which enabled to describe three aspects of the Israeli fourth grade mathematics education: the intended curriculum, the implemented curriculum, and the attained curriculum.

As to the intended curriculum, the National Syllabus, as well as three widely used mathematics instructional programs, were analysed. It was found that 22 different mathematics instructional materials sets (instructional programs) were in use in the country, and a substantial proportion of these sets has not been approved by the Ministry of Education and Culture. Three instructional programs have had a wide dissemination in the country. These three programs strongly emphasize the goals specified in the 1972 National Syllabus, and they differ from each other mainly with regard to selecting the learner's natural environment as a basis of teaching mathematical concepts and skills versus creating an artificial environment in the class which leads to the comprehension of the concepts dealt with in the textbooks. Another difference is the degree of formal definitions of concepts studied and the role of these definitions in the framework of the classroom work.

The implementation was examined on the basis of its seven salient components: continuity, using the syllabus, using the Teacher's Guide, having teachers' meetings dealing with mathematics instruction, instructional guidance, teaching aids and class organization.

As to the attained curriculum, in contrary to expectations no differences were found between the achievement levels on word problems and computational skills. On the other hand, students performed better on test items which dealt with mathematical topics already introduced in grade three, than on topics introduced only at grade four. The set of 40 items contained a subset of 10 items which represented Minimum Competency Requirements. It was found that approximately 5 percent of the students failed to attain mastery on the Minimum Competency test imbedded in the 40 items achievement test. In approximately 15 percent of the schools there were more than three pupils who failed to master the Minimum Competency Requirements. These schools constitute a high level of hazard, and priority should be given to them in the allocation of compensatory education budget.

List of Tables

Number	Content	Page
2.1	Schools according to the number of parallel fourth grade classes in the whole Israeli school system	47
2.2	A sample of 300 schools	48
2.3	The sample of 100 schools	50
2.4	Number of items in various behaviors by content categories	55
2.5	Four schemes of item classification	57
2.6	Items in the two Minimum Requirement item series	58
2.7	The frequency distribution of scores on the series of minimum tests	60
3.1	Instructional programs used in classes	66
4.1	Father's education according to the compensatory index of schools (percent)	79
4.2	Utilization of calculators at home and in the school	82
4.3	Teacher's basic professional education level	84
4.4	Teacher's professional education in mathematics	85
4.5	A schematic representation of Eraut's (1972) four in-service education paradigms	86
4.6	Teacher's autonomy in deciding adopted instructional program	89
4.7	Teacher's attitude towards the textbook used in teaching mathematics	93
4.8	Allocated time in fourth grade mathematics	96
4.9	Teacher's actual implementation behaviors	101
4.10	Statistics of the OTL scores	108
4.11	Frequency distribution of the two OTL scores	110
4.12	Opportunity to learn each of the 21 curricular topics	111
4.13	Distribution of schools across levels of coverage of curricular topics	112

Number	Content	Page
4.14	Distribution of schools across level of coverage of the test items	113
5.1	Means and standard deviations obtained on the total test and on subsections of it	117
5.2	Frequency distribution of the total test percentage scores and of the minimum competency percentage scores	119
5.3	Homogeneity of achievements across schools in fourteen countries	122
5.4	Distribution of low-achievers across classes	124
5.5	Number of schools with different levels of success in pairs of items	132
5.6	Percentage of test items difficulty at the whole system level and at class level	135
5.7	Attitude toward mathematics	150
5.8	Distribution of the attitude scores	151
6.1	Cognitive achievements in six school districts	153
6.2	Mathematics achievement levels of boys and girls	156
6.3	Distribution of girls in the classes	157
6.4	Distribution of achievement differences of boys and girls on the total test	157
6.5	The relationship between sex composition of classes and sex-related achievement differences over classes	158
6.6	Cognitive achievements of three priority groups	161
6.7	Cognitive achievements in three program groups	162

List of Figures and Graphs

Figures

Number	Content	Page
1.1	Tyler's Model of Evaluation	14
1.2	Davis' (1981) revised version of Lewy's (1977) mapping sentence definition of evaluation	16
5.1	Distribution of fourth-grade mathematics scores in two schools	122

Graphs

Number	Content	Page
5.1	Achievement profiles on seven subscales of the whole population	140
5.2	The profiles of three textbook groups	142
5.3	Achievement profiles on seven subscales of two schools	144
5.4	Achievement profiles (scores in percentage) on seven subscales of two pupils	146
5.5	Two pupils' T achievement score profiles calculated in the form of deviation from the population mean on seven subscales	147
5.6	Two pupils' T achievement score profiles calculated in the form of deviation from class means on seven subscales	148

List of Exhibits

Number	Content	Page
1.1	An example of NAEP report's profile	10
1.2	An example of Pennsylvania report's profile ..	11
1.3	Mathematics achievement progress of two ethnic groups in grades 4 and 6 (Lewy and Chen, 1977)	13
1.4	Excerpts from the grade four syllabus 1982 ...	38
1.5	Selected items from the parenthesis program ..	38
3.1	Grade four teaching contents listed in the 1972's national syllabus	67
4.1	Contents of two item which were covered in all schools	114
4.2	Contents of two items which were covered in few schools (less than 60% of the schools) ...	114
5.1	Content of item no 2	129
5.2	Content of item no 4	129
5.3	Content of item no 25	130
5.4	A hypothetical example of two achievement profiles	138

Chapter One
The Theoretical Framework

The Background of the Study

In the last three decades one can observe an increasing demand for systematic information both on the effectiveness of various educational programs and on the operation of school systems and subsystems. The demands for examining educational programs have been supported by the ideas of Campbell (1969) about the "experimenting society", which implied that social amelioration programs should be tried out before they are put into use. Boruch (1982) put even stronger demands and claimed that in determining the effectiveness of such programs one should prefer rigorously planned, randomized experiments. Simultaneously the call for an examination of the effectiveness of school and program management was echoed by the rapidly increasing demands of accountability in the 1970's (Browder, 1985). In Israel, too, one observes an increasing demand for systematically collected and professionally processed and interpreted information about various aspects of the educational system, and about the quality of programs and materials used by the system. In response to this demand, and since the establishment of the Israel Curriculum Center, most new curriculum materials have been evaluated before being put into use in the whole educational system, and evaluation became a permanent component of the development process (Lewy, 1971).

At the same time, attempts have been made in Israel to examine outcomes of education, and several national assessment projects have been carried out (Lewy and Chen, 1977; Minkovits et al, 1977; Razel et al, 1984). In 1979, the chief scientist of the Ministry of Education and Culture appointed a Committee to examine the necessity and feasibility of establishing a Feedback System that will continuously provide information for decision makers at various levels (national, district, school, class, individual student) of the educational system. The committee was also

asked to propose a framework for the operational aspects of the Feedback System, such as determining information needs, providing information, controlling access to the information, etc. The recommendations of the Committee were released to the general public two years later (The Henrietta Szold Institute, 1981).

The Committee recommended giving high priority to data about cognitive achievements at various grade levels. Up to the time of the present study, no formal measures have been taken by the Ministry of Education to establish a Feedback System. Two operational units within the Ministry did start, however, to collect and to process information about the attainment of cognitive educational objectives following a pattern which may serve as a model for a Feedback System in Israel and in other countries across the world. The Office of the Chief Scientist commissioned a series of studies for examining the attainment of Minimum Achievement Requirements in reading and mathematics at grade three of the Elementary School. Within the framework of these studies, all grade three children in the system (approximately 70.000 pupils) were tested. The aim of these projects was to identify pupils who are in need of remedial teaching and thus to minimize the number of children who might not achieve a minimum level of literacy and numeracy in the Elementary School (Razel et al, 1984).

The second unit within the Ministry of Education and Culture, which has carried out comprehensive evaluation studies pursuant to the recommendations of the Committee, is the Israel Curriculum Center (ICC). In cooperation with the Pedagogical Secretariat, a plan was worked out wherein a series of studies will be carried out in a cycle of six years, with the aim of providing a comprehensive picture of the state of the art of teaching, and of the outcomes of teaching in a particular subject area. Studies have so far been carried out in the following areas: History (Wolff, 1983), Science (Lewy et al, 1983), Bible Studies (Padua, 1986), Geography (Raz, 1986), and studies in Writing and in English as a Foreign Language are in process at present. These studies followed the principles of a multipurpose feedback system outlined by Lewy (1979). Accordingly, they may serve, simultaneously, as a source for a variety of pre-specified or emerging information needs such as: changing curricula, allocating budget, changing the method of teaching, or providing guidance for a particular group of schools, etc.

Within the framework of this series, the Israel Curriculum

Center carried out a study on the state of the art of teaching-learning mathemathics in the fourth grade of elementary school. This report constitutes the first comprehensive summary of the data collected. Being a multipurpose summary it sheds light on the teaching-learning conditions of mathematics in school, on the overall attainment level of the fourth-grade elementary school pupils in Israel, on the proportion of children who mastered basic numeracy skills, and on the relative effectiveness of various alternative curricula[1].

Beyond the attempt to answer specific questions related to the state of the art of mathemathics teaching-learning in the fourth grade of the elementary school, the present study also served to explore the utility and feasibility of carrying out multipurpose assessment studies within the framework of a National Feedback System. In this respect, the present study constitutes a continuation of a series of studies initiated by the Committee appointed by the Chief Scientist of the Ministry of Education and Culture.

When the present study reached a stage of near completion, the Ministry of Education re-vitalized the plans for establishing a Feedback System on educational achievement and made steps for starting its operation. The present study, which was planned when the matter of the Feedback System was in a dormant phase, may serve as a timely response to questions the Israeli Feedback System may encounter at the start of its operation. Moreover, the present study may serve as a model for multipurpose evaluation studies in other school systems which intend to establish Feedback

[1]It should be noted that the Israel Curriculum Center released a short summary report of the findings for the use of the teachers whose classes participated in the study (Lewy and Raz, 1985). A more detailed summary of the study is at present in the process of preparation by a team of the Curriculum Center in which the writer of this research report also participated. The writer of this report participated in the work of the curriculum team, which planned the study and fulfilled a leading role in developing the measurement and data collection instruments, and in determining the procedures for extracting the sample. The data analyses reported in this study and the interpretation of the data was done partly parallel with, but fully independently from that carried out by the ICC team. It reflects the views and interpretation of the writer of the report and in no way do they represent the views of the Ministry of Education and Culture.

Systems. In the last decade, the interest in establishing Feedback Systems has been strongly felt in South American countries as well. The participation of several South American countries in an extension study of the International Association for the Evaluation of Educational Achievement on outcomes of mathematics education has stimulated an interest in these countries for conducting national studies on the topic. Thus, the Israeli mathematics assessment study may serve as a useful model for South American countries, among them Brazil, the country from which the writer of this report came to study in Israel.

From a theoretical point of view, this study touches on three major concerns extensively discussed in the educational literature of the last decade: the role of assessment and evaluation in educational planning; new approaches to teaching elementary school mathematics; and the concept of Minimum Achievement Requirement in the field of literacy and numeracy.

The forthcoming sections of this chapter will spell out basic issues related to these concerns.

Evaluation and Assessment

Changing Conceptions about Examining Educational Achievements

For centuries, the examination of outcomes, which was assigned to teachers, was considered to be a relatively marginal task in education. In most cases, teachers conducted oral examinations, asking their students to respond to questions directly related to contents taught in school courses. Frequently, examinations, and mainly end-of-year-examinations, constituted a ritual in which parents, representatives of the community and school authorities were allowed to participate, and to also pose questions to the examinees. Conducting examinations was not considered to be a complex component of school life, neither was it considered to be a professional skill requiring special training. For a very long time, the right of teachers to set examination questions was fully aknowledged and nobody cast any doubt about the adequacy of this procedure.

The first attempts to utilizing scientific principles for

conducting school examinations was made by the French psychologist Alfred Binet at the end of the 19th century. Binet developed tests for measuring the intelligence of school children and by doing this, he began a movement which led to the emergence of testing as a professional activity in its own right. Testing gained recognition mainly in the United States and experts in testing started to operate inside schools, measuring the level of the educational achievement of learners. The skills of developing tests, administering them, and interpreting their results became fully separated from the skills associated with the teaching profession. Binet was a psychologist, and those who developed the first tests for measuring educational achievements of children also came from the profession of psychology. Constructing tests and using tests became a specialized branch of psychology and later of educational psychology.

Its orientation was quantitative. The utilization of tests in schools became labelled as educational measurement and the scientific principles governing its utilization came to be known by the name psychometrics. For decades, educational psychologists referred to the utilization of tests as measurement thereby alluding to the scientific nature of their endeavor and the precision of their results. Only in the late 1930's did a new term for labelling activities appear, which aimed to examine educational achievement. The Eight Year Study in the United States, which thoroughly examined the impact of the Progressive Education movement (Smith and Tyler, 1942), referred to activities examining educational achievement as evaluation. The term evaluation was used to describe both the attainment of operationally-defined educational objectives and changes in the behavior of the learner. In his well-known evaluation model, Tyler (1949) defined three major foci of education, one them being evaluation.

In the 1940's a new expression was added to the terms describing activities which were concerned with examining human traits and among them educational achievements. Stern (1956) introduced the term assessment. This term was used for labelling multiple activities carried out by a team of psychologists with the aim of describing the characteristics of a person, among them the person's educational achievements. The psychologists observed the spontaneous behavior of army recruits staying for several weeks in a military camp. They carried out interviews with them, administered psychological and achievement tests, and on the basis of integrating the information gained from various

sources, the team of psychologists attempted to assess a person's capacity to fulfill highly dangerous special tasks in the army. The psychologists used multiple methods for assessing the characteristics of a person and practiced an approach which later became labelled as triangulation (Denzin, 1978). Since then, the three terms of measurement, evaluation, and assessment have been extensively used in the educational literature. The first one of these terms has been defined as: highly structured activities aiming to yield numerical measures of human traits. Experts working in this field strive to increase the precision of these measures, and for attaining this goal they developed highly sophisticated mathematical theories of measurement.

The second term, evaluation, has a broader meaning and it is used mainly to examine the impact of a particular action. Accordingly, evaluating the educational achievement of a person, or of a group of persons, is used to describe activities which determine whether desired changes occurred in the level of their educational achievement. More frequently, however, "evaluation" designates activities which aim to determine the impact of actions, programs, and projects. In an educational context, an evaluation study may examine the impact of a certain educational intervention (using a particular teaching style, instructional materials, resources, organizational arrangements). The criterion for measuring impact may be changes in the educational achievements of the learners, but the target of the evaluation is the intervention and not the educational achievement of those who were tested within the framework of the study (Choppin, 1985). It should be noted that, in practice, we frequently encounter a vague use of the term evaluation, but theoristis like Tyler (1949) or Choppin (1985) strictly adhered to this distinction.

There is a greater ambiguity with regard to the utilization of the third term, assessment. The meanig attributed to this term by Stern (1956) has not been accepted in educational literature. Recently, Choppin (1985) suggested reserving this term for describing activities which focus on describing the scholastic achievements of individuals. Accepting Choppin's (1985) suggestion would mean that conducting examinations or testing for the sake of determining scores of individuals, and assigning school marks to individuals would be labelled assessment. But Choppin, himself, is aware of the fact that this definition of the term is not accepted in practice. He writes:

Note that the large monitoring program within the

> *United States, the National Assessment of Educational Progress (NAEP), is not an assessment within this definition. Also, individual students are given tests, no interest is attached to their individual results....NAEP is an example of evaluation rather than assessment.* (p. 1747)

The ambiguity of the term assessment is well illustrated by the note of Choppin.

Due to this ambiguity, in our report we will use the term assessment as it is commonly used in the U.S., i.e. a study which primarily focusses on examining educational achievement, like the NAEP study, without attempting to relate variations in achievement to differences in treatment. On the other hand, we will use the expression evaluation to studies in which the measure of achievement is used as a criterion variable for comparing the impact or the relative effectiveness of a particular educational treatment. Only casual reference to the term measurement will be made in our study, and problems of measurement will be dealt with only marginally in this study.

Trends in Assessment

As indicated above, by assessment we refer to examination of educational achievement in a particular population without reference to factors which explain the variance in achievement levels. An assessment study may provide individual scores for all children in a well-defined population, as in the assessment studies of grade three pupils in Israel, carried out by the Szold Institute (Razel et al, 1984), or alternatively, it may provide statistics about a particular population or a well-defined substratum of a population on the basis of data collected from a sample, such as the National Assessment of Educational Progress (NAEP) in the U.S. (Greenbaum, Garet and Solomon, 1977). The purpose of the NAEP study is to furnish information for the public and particularly for those having policy responsibility for educational achievement in various sectors of the American learners-population. The project operates as an independent research institute, and the Federal Government provides funds to finance its activities. The institute carries out continuous testing in most subject ares taught in schools (reading, writing, mathematics, science, citizenship, social studies, literature, art, music, and career and occupational development). The

assessment is conducted for the following age groups: 9, 13 and 17 year-olds and for young adults. The results are reported separately for the following subgroups: geographic regions, type of community (rural, low metropolitan, high metropolitan), size of the community, parental education, sex and race.

An entirely different approach to assessment is employed in the United Kingdom by the project known as the Assessment of Performance Unit (APU) of the Departament of Education and Science (DES). The goals of the British project of assessment were defined as "*...assessing the effectiveness of the educational system and monitoring pupils' performance across the curriculum at different ages*" (Pring, 1981, p. 156). The APU does not provide individual scores, nor does it publish statistics about particular schools but it does provide a base from which individuals look at their own performance. It also may help educational planners with ideas about issues which need to be thoroughly studied. Unlike the American assessment study, which is structured according to school disciplines, the British study specified six areas or lines of development, which transcended particular subject boundaries. These lines were: mathematical, language, scientific, physical, aesthetic, and personal and social. Later it was decided to discontinue the monitoring of the "personal and social" line, and instead a modern language (foreign) line was incorporated into the project. Pring (1981) points out several unresolved problems of the APU which equally applies to other national assessment studies.

(1) The adoption of a particular curriculum model - The unit has defined six lines of development, but the selection of six lines or broad areas of studies is arbitrary, and unconvincing.

(2) Focus on lines of development - Due to the diversity of syllabuses used in various schools the assessment focuses on lines of development rather on specific curricular contents. This decision is based on the assumption that one may identify common lines of development across various syllabuses and curricula, and thus it is possible to develop tests which are not biased toward a particular textbook. This assumption has not received empirical support and in a large scale study it was found that very few tests remain unbiased to a particular textbook (Walker and Schafferzik, 1974).

(3) The selection of objectives - The APU aimed to provide a picture of pupil performance. Its mandate was to describe

achievement but not to influence what should be taught in school. The planners of the project wanted to avoid schools aligning themselves to the content of the tests. They strived, rather, to produce tests which reflect what is actually taught in school. Therefore, information had to be gathered about what is actually taught, and only those objectives commonly taught in most schools were considered for testing.

Pring (1981) claims that this approach to test development entails several problems. First, the teachers' description of what is taught in the school does not always reflect the reality:

> *The connection between what in fact happens and what teachers say is happening are often tenuous. This is due due partly to the elasticity of the concepts that people employ to describe teaching different kinds of development (people use the same words to describe different events, and different words to describe the same event) and partly the lack of objectivity in most accounts of school and classroom experience. There is not an unambiguous and universally aggreed language of educational practice* (p. 169).

Parallel to the conceptual differences in setting up national assessment projects one may observe a great variation in ways the results of the assessment studies are reported. The Committee for Examining the Utility and the Feasibility of a Feedback System for the Ministry of Education and Culture (The Henrietta Szold Institute, 1981) provided details about reporting patterns of several large scale assessment projects.

The NAEP reports the results separately for each item included in the test, as it is presented in Exhibit 1.1. The same approach is employed in the Israeli Assessment Study of Achievement in the Arabic School System of the country (Bashi et al, 1981).

A different approach is employed in the Pensylvania assessment study (Pensylvania Department of Education, 1973). The Pensylvania Assessment study provides results separately for each school on thirteen major educational objectives examined within each of the assessment studies. Each school receives a one-sheet summary which specifies the percentile rank of the school on each of the objectives.

In Exhibit 1.2 the percentile rank location of a particular

A profile of an item (NAEP study report)

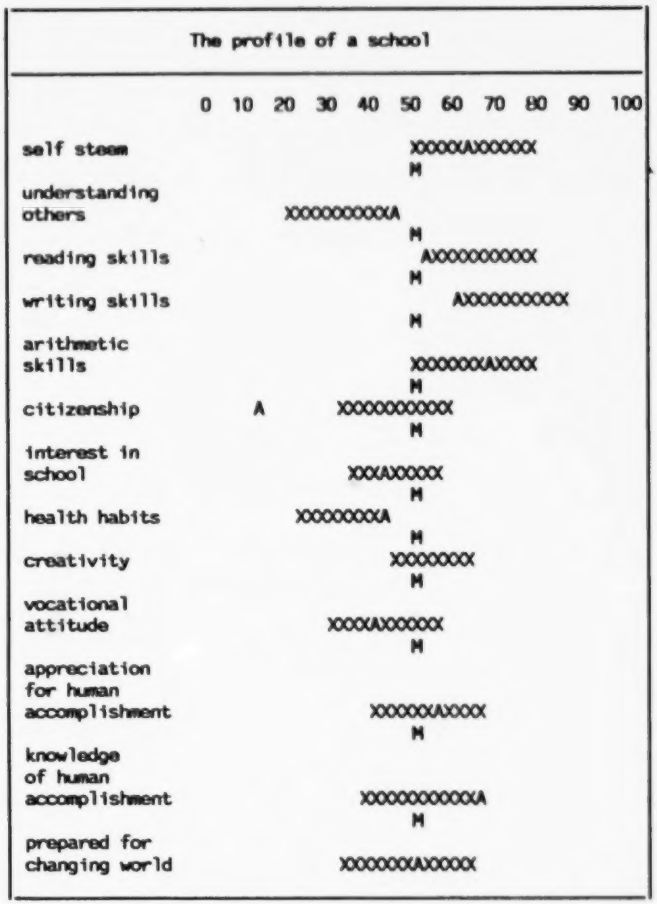
Percentage of Success. Starred Effects Indicate Reliable Differences

% of Success	Group Effect	Reporting Category	Group	0 10 20 30 40 50 60 70 80 * * * * * * * * *
28.2	- 3.3	National		*********
		Region		
24.0	- 3.3		Southeast	********
19.8	- 8.4		West	*******
29.8	0.8		Central	**********
36.9	8.8		Northwest	**********
		Sex		
19.3	- 8.9		Male	******
38.5	10.3		Female	**********
		Color		
12.3	-15.9		Black	****
31.6	3.4		White	**********
		PARENTAL EDUCATION		
21.9	- 6.4		No High School	**********
33.7	5.5		Some High School	**********
23.8	- 4.4		Graduate High School	**********
38.3	10.1		Post High School	**********
		SIZE AND TYPE OF COMMUNITY		
10.2	-18.0		Extreme Inner City	*****
14.0	-14.2		Extreme Rural	******
28.4	0.2		Small City	**********
31.6	3.4		Medium City	**********
19.2	- 9.0		Rest of City	********
31.4	3.2		Suburban Fringe	**********
37.4	9.2		Extreme Aff Suburb	**********

Exhibit 1.1 An example of NAEP report's profile

school is marked by the letter of A. An array of Xs specify the predicted range position of schools of similar descriptive characteristics (such as socio-economic level, entry level test score, etc.). The reader may observe that in creativity the

school scored above its expected range, and in citizenship below the range. In other areas, its scores were within the range of expectations. The recipient of such a summary can easily observe whether a particular school's actual attainments are in the range of the expected, or whether it is below or above it.



Exihibit 1.2 An Example of Pensylvania report's profile

Such information serves as a diagnostic tool for a school and it may serve as a basis for decision makers within the school to ascertain which particular study areas need to be strengthened or

improved in the school.

Large scale assessment studies were carried out in Israel in major subject areas mainly in the elementary school. For 19 years (during 1955-73) the Ministry of Education and Culture conducted an annual survey of educational achievements in grade eight (The SEKER examinations). The results of these examinations served as a screening device for the publicly supported Academic Secondary Schools[2] (Orthar, 1967). Another study, a longitudinal one, was carried out by Lewy and Chen (1977), with the aim of assessing whether the achievement gap between children of immigrants from developed countries and of those from developing countries remains constant during the course of their studies in the school or alternatively they increase during the years.

The graphic representation of the progress of the two ethnic groups during their studies in grade four and six and the interpretation of the results provided by Lewy and Chen (1977) appear in Exhibit 1.3.

The achievement gap between children from various ethnic groups was the major concern of the Van Leer study (Minkovits et al, 1977). The Van Leer study examined differences between the achievement of children from the two ethnic groups whose parents immigrated to Israel and those whose parents were already born in Israel (i.e. their grandparents or even an earlier generation immigrated to the country), and found that the gap between the ethnic groups of third generation immigrants is smaller than the gap between the second generation immigrants. The authors concluded that the achievement gap between various ethnic groups diminishes gradually from one generation to another.

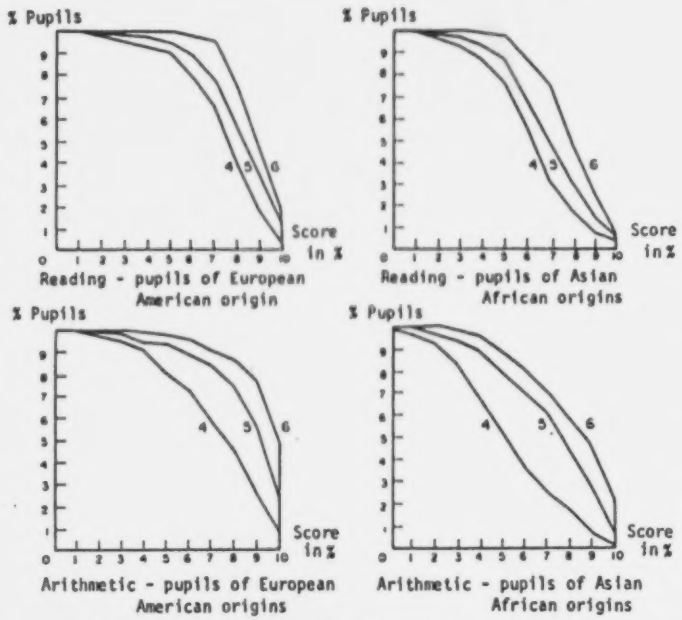
Finally, to a series of national assessment studies in various subject areas which were carried out by the Israel Curriculum Center reference was made previously. The present study constitutes one of that series.

Trends in Evaluation Studies

As indicated above, by the term evaluation we refer to studies

---

[2]Those who failed the SEKER examination were admitted only to the Vocational Secondary Schools, or could opt going to private schools.


% Pupils
Score in %
4
5
6
Reading - pupils of European American origin
Reading - pupils of Asian African origins
Arithmetic - pupils of European American origins
Arithmetic - pupils of Asian African origins

Cumulative frequency distribution of anchor items for the two ethnic groups (grade levels are indicated next to graph lines).

In order to compare the academic progress shown by students belonging to the two ethnic groups, cumulative frequency graphs were drawn separately for each of the two groups from their scores on the arithmetic, reading, and English tests. Scanning the graphs in Fig. 6.2, one notices that the typical form taken by the progress curves of the European/ American pupils is convex, while the characteristic curve for the African/Asian students is concave. In other words, there is a concentration of students from African and Asian origins around the lower scores and a clustering of students from European and American origins around the higher scores.

Exhibit 1.3 Mathematics Achievement progress of two ethnic groups in grades 4 and 6 (Lévy and Chen, 1977).

which try to determine the effectiveness of a particular treatment. In an educational context, the treatment is usually a particular program, curriculum set of instructional materials, organizational arrangement, etc, and the criterion of its effectiveness is the mastery of specified educational objectives, or the attainment of desired changes in the behavior of the learner. Scriven (1967) defines evaluation as an act of determining the worth of a certain entity. He qualifies this definition by appending to it the note that the concept of

evaluation can be applied to entities of any type, and whether the entity is an educational program or a coffee machine, the process of evaluation remains the same. It consists of posing types of questions about the entity, such as: "How well does it perform?", and "Is the use....worth what it is costing?". Answers to these questions are usually provided by summarizing information related to cognitive or affective achievements of the learner. It should be noted that such achievements are by no means the sole criteria for determining the effectiveness of an educational program. The criteria of effectiveness may touch on strengthening the school-community relationship, or increasing the involvement of parents in school affairs, and so on, but since the major goals of schooling are to change the learners' behavior in a desired direction, the ultimate criteria of effectiveness is always related, directly or non-directly, to students' achievements. Hence the strong association between the concept of evaluation and testing educational achievements. Conducting testing, or measuring educational achievement, is by no means the focus of the evaluation. The object of evaluation is the intervention.

Lewy (1977b) summarized the state of the art of curriculum evaluation in the middle of the 1970's. He discerned three competing approaches domminant at the time. These were concerned with: (1) achievement of desired outcomes; (2) assessment of merit; and (3) decision making.

(1) Achievement of desired outcomes - This approach was best represented through Tyler's (1949) model, as presented schematically in Figure 1.1.

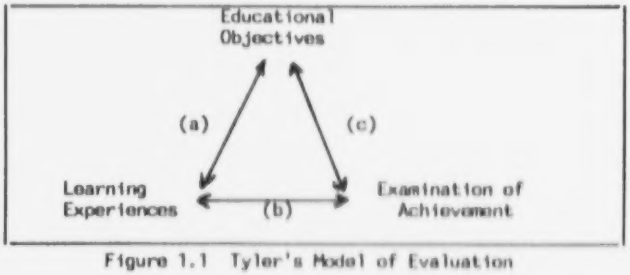


Figure 1.1 Tyler's Model of Evaluation

The evaluator working with this model is interested to know whether students mastered the objectives. In Figure 1.1 this type of evaluation is represented by the arrow marked

with the letter (c). Tyler's model was criticized as restrictive in that it does not deal with occurrences of unplanned or unintended events. Stake (1969) has raised the criticism that this model unduly emphasizes the outcomes of the program and does not pay attention to process variables or to the examination of antecedent conditions that affect the success of the program. Criticism of this type led to the emergence of alternative models of curriculum evaluation.

(2) The merit of an entity - The definition of evaluation as an act of determining the merit of an entity has been mentioned earlier.

(3) The decision oriented approach - The decision oriented approach is best epitomized by Alkin's (1969) definition of evaluation.

> *Evaluation is the process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analysing information in order to report summary data useful to decision-markers in selecting among alternatives.* (p. 2)

Based on an examination of empirical studies of curriculum evaluation in the 1970's, Lewy (1977b) suggested a mapping sentence definition of evaluation which contained six facets.

A slightly revised version of this definition is presented in Figure 1.2.

Ten years after the first publication of the mapping sentence definition of curriculum evaluation, one may ask to what extent do the six facets represent the major concerns of evaluation experts in the second half of the 1980's.

Nevo (1983) summarized the major concerns of curriculum evaluation by compiling a list of persistent dilemmas. Some of these dilemmas are already represented in the previously mentioned mapping sentence. Thus, for example, he mentions the facet of roles, the object of evaluation (labeled in the mapping sentence as entity), methods of enquiry, etc.

Examining the state of the art of curriculum evaluation in the

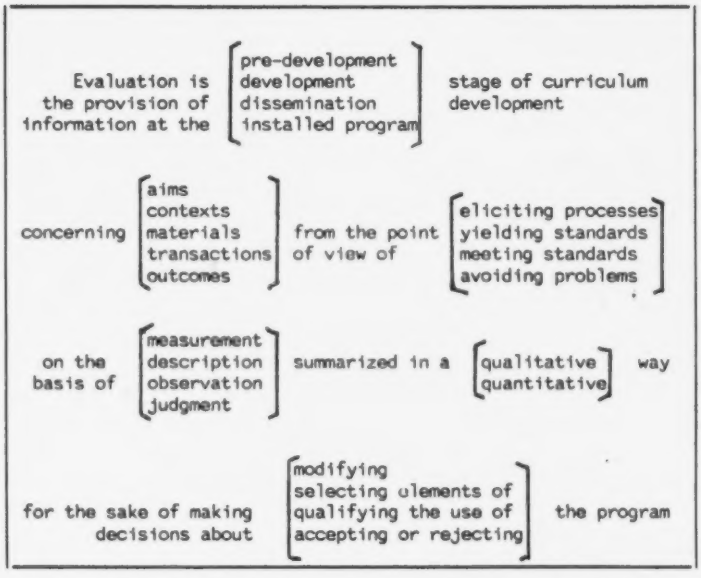


Figure 1.2 Davis' (1981) revised version of Lewy's (1977b) mapping sentence definition of evaluation.

middle of the 1980's with what prevailed 10-15 years earlier, some trends are discernible which were partly forecast in Lewy's(1977b) mapping sentence definition. They are related to the following facets of the definition: stages, data, mode of summary, and role.

Stages - The 1977 definition mentions implementation as one of the stages at which evaluation should be carried out. Interest in evaluating curriculum evaluation has grown since 1977, and both the theoretical and the practical aspects of implementation became highly elaborated. Implementation gained a clearer conceptualization and the technology of evaluating it has been greatly improved (Fullan, 1981). As a matter of fact, it has become a target of evaluation, and one may list implementation also as an element of the entity facet.

Data - With regard to data, there is a shift toward increased utilization of observational data collected in naturalistic settings. Interest has also been revealed in archival data, and in unobstrusive data of other types

(Luckey et al, 1983), in retrospective data measuring long-range impacts of curricula (Tamir and Amir,1981), and in photographic records of past events (Templin, 1979).

Mode of Summary - The mapping sentence definition from 1977 specified two alternative ways of data summary: quantitative and qualitative. At that time the qualitative data summary was seldom used. Over the last ten years its importance gradually increased and it became the dominant mode of data summary in curriculum evaluation.

For several decades the quantitative and qualitative approaches were considered conflicting views and the gulf between them has constantly deepened. This has happened despite the fact that more than a quarter century ago Zetterberg (1962) had already advanced the hypothesis that quantitative and qualitative approaches have distinct roles in the advancement of disciplinary knowledge, and both of them are essential. Qualitative field-work can help to illuminate the theoretical and measurement issues in new areas of study, while quantitative methods are appropriate for testing the hypothesis derived from rigorous field methods. By the end of 1977 one may already observe a kind of reconciliation, and Jick (1979) suggested the need to mix the two methods.

The role of evaluation - The mapping sentence definition specifies several alternative roles evaluation may fulfill in the process of curriculum development. In the last ten years we witnessed a trend of broadening the scope of roles evaluation may fulfill. Alkin et al (1979) suggest ways which may ensure evaluation results being used so that evaluation may effectively fulfill the roles attributed to it. Recently, Couisins and Leithwood (1986) surveyed and systematized the findings of approximately 70 empirical studies which examined issues related to evaluation utilization.

Finally, two other trends in evaluation should be mentioned here because they had great impact on the study described in this report: sensitivity to the needs of the evaluation clients, and the utilization of multiple methods in examining a single evaluation target.

The needs of the client - Alkin et al (1979) emphasized the importance of considering the information needs of clients. They indicated that various groups may be interested in obtainning information about the effectiveness of a particular entity, and in most cases it may be useful to produce several evaluation reports to satisfy the differential needs of various groups. This approach was followed in the present study, too. The evaluation team conducting this study had an interest in reaching various client groups, and as described in a later section of this report, they prepared several evaluation reports, each one geared towards the needs of one particular audience group.

Utilization of multiple data sources and methods - The rationale for using multiple methods is not new in the field of curriculum evaluation. Atkin (1979) wrote about efforts to evaluate the status of science teaching in the United States, and described an attempt to simultaneously carry out three studies, each one employing a different method. These three methods were: conducting a series of case studies in several schools; a survey study based on data collection through questionnaires; and an analytical review of the professional literature on this topic. He concluded that

> *...thus by triangulating on science teching, using three differenct approaches, it was hoped that a reasonably accurate picture may be developed as a benchmark of the current state of science in the schools* (p. 397).

Developments related to utilizing multiple data sources and methods in a single study also had great impact on determining the design of the present study.

Links Between Evaluation and Assessment

As already indicated, both evaluation and assessment studies reflect an interest in the educationl achievement of learners, even though the target of evaluation is a particular educational intervention, while the target of assessment studies is the achievement of the population examined in the study, without directly relating variation in achievement to variation in treatment. During the last thirty years a great many evaluation and assessment studies were carried out across the world, and the increase in the volume of such studies went together with increased differentiation between the conceptual bases and

methodological features of these two types of studies. As Cronbach (1975) wrote about the two psychologies, one may also observe a gulf between two evaluation traditions, with little interaction between them. The lack of interaction between these two patterns of study was frequently noted and deplored by researchers. Cunningham (1976) claimed that assessment studies should not disregard the importance of identifying and examining the causes of achievement differences observed in an educational system. He stated that assessment enterprises "*...must take on the added dimmension of being an aid for policy formation and decision making*" (p. 85), and that assessment has "*...to include input policy variables related to educational attainment*" (p. 87).

This idea appears also in the early writings of Stufflebeam, too. In the article "Evaluation as Enlightment for Decision Making" (Stufflebeam, 1969) he observed that the objective of product evaluation (or evaluation of learners' attainments) is to measure and interpret outcomes by relating learners' achievements to objectives and to context, input, and process evaluations. Moreover, he observed that the method of doing product evaluation involves, in its ultimate phase, "*...rational interpretations of the outcomes using the recorded context, input and process information*" (p. 65).

The present study constitutes an attempt to strenghten the links between evaluation and assessment. It is rooted in the belief that one should reduce further differentiation between these two types of disciplinary endeavors, and that both evaluation and assessment studies will better fulfill their designated roles if the evolution of the two separate "evaluation" discipline is avoided. To attain this goal, the study strives to demonstrate the feasibility and the utility of using a single set of data, within the framework of a single study, to fulfill roles which hitherto have been attributed either to evaluation studies or to assessment studies, which have been carried out without creating an integrative link between them.

Trends in Elementary School Mathematics Curricula

Since the topic of our investigation is fourth grade elementary school mathematics in Israel, and we are interested to examine both conditions and achievements in this field, it seems necessary to provide some information about world trends which affected

mathematics teaching in Israel in the last three decades.

As indicated by Kliebard (1985), in the 1950's a new curriculum revision movement emerged in the United States, which changed the content and the method of teaching of most subjects in the U.S., as well as in other countries of the world. Kliebard indicates that this movement had already begun in 1951, six years before the Soviet Union launched its Sputinik:

> *As early as 1951, the University of Illinois Committee on School Mathematics began to consider the effectiveness of mathematics teaching in elementary and secondary schools and, a few years later, Jerrold Zakarias, a physicist at the Massachusetts Institute of Technology organized an effort to reconstruct secondary school physical sciences. Momentum in curriculum matters began to shift from professional educators to academic mathematicians and scientists, and from curricula that emphasized direct utility in everyday affairs to rigorously academic and highly structured programs of study.* (p. 1228)

Curriculum development in the field of mathematics was already very intensive at the early phases of this movement. It was felt that mathematics, as taught in schools, failed to give a sufficiently sound understanding of the subject to meet the needs of an increasingly technological age (Taylor, 1976). Levin (1985) lists two reasons why earlier school mathemathics programs failed to achieve their goals:

(1) The programs neglected the meaning and rational principles which make arithmetic a significant phase of mathematics.
(2) The programs gave students little chance to use previously learned ideas and skills in solving their own personal problems. (p. 315)

It was also felt that a substantial proportion of children do not like mathematics and that its study was a heavy burden. The developers of new mathematics programs assumed responsibility for producing new curricula which would help the pupil both to understand and to enjoy mathematics.

The new mathematics programs emphasized understanding, originality, and creativity, and put less emphasis on routine calculations.

Further, it was felt that giving children an intellectual challenge would improve their understanding of and attitude toward mathematics. Therefore, as stated by Resnick and Ford (1981):

> *...in schools, youngs were exposed to relatively advanced concepts like inequalities, the properties of sets, the use of zero as a number, and the principles underlying the decimal notation.* (p. 101)

The comparison of the new textbooks for mathematics (produced in the wake of the curriculum movement) with earlier books provided evidence that the changes affected the secondary school curricula more than that of the elementary schools. In secondary schools, both method and the content of teaching were modified. While in elementary school textbooks one may also observe considerable changes in the content, major changes occurred mainly in the methods of teaching mathematics, and in the sequence of teaching certain content elements. The difficulties young children encountered in acquiring basic math concepts imposed a higher level of responsability on primary school mathematics (Taylor, 1976). Therefore at the primary school level it was necessary to consider the limitations of the child's ability to absorb complex ideas, to deal with abstract concepts, and to carry out formal operations.

The new mathematics movement, to a large extent, adopted the Piagetian theories of child development and could not disregard its implications for constructing mathematics curricula. The Piagetian theory of child development constituted a basis for mathematics curricula prior to the advent of the new curriculum movement in the 1950's in the United States, Quite likely it enlightened textbook writers, and teachers intuitively incorporated some Piagetian ideas in their pedagogical practice long before Piaget's theories were explicitly formulated. This adherence to Piagetian ideas in constructing primary school mathematics curricula served as a stabilizing factor, but it did not fully hinder the emergence of other innovations in primary school mathematics. This happened for at least two reasons. First, the interpretation of the implications of Piaget's theories for curriculum construction underwent changes, and secondly, new theories relevant to the acquisition of mathematical knowledge emerged and became adopted by curriculum developers.

In reviewing how the implications of Piagetian theory have been variously interpreted for curriculum development, we find that the

ideas of Bruner are most frequently quoted.

Bruner (1960) argued against the conception that a readiness to learn something is a matter of maturation, and one has to postpone teaching certain skills until the necessary maturation has occurred. He was against the practice of teaching children pre-mathematical exercises, all the while waiting for that readiness to develop. One of his most quoted statements, which had an impact on the improvement of new curriculum development, is:

> *We begin with the hypothesis that any subject can be taught effectively in some intellectually honest form to any child at any stage of development... The task of teaching a subject at any particular age is one of representing the structure of that subject in terms of the child's way of viewing things....Thus instruction ...even at the elementary level, need not follow slavishly the natural course of cognitive development in the child. It can also lead intellectual development by providing challenging but usable opportunities for the child to forge ahead in his development.* (p. 5)

Bruner's ideas motivated mathematicians to develop curricula which teach mathematics as a structured discipline. This meant that mathematics was not expected to be taught as a random collection of procedures for carrying out computations, but rather as "*...a unified system of concepts and operations that explain certain patterns and relationships that exist in the world*" (Resnick and Ford, 1981, p. 105). These authors provided an example how mathematics can be taught as a structured body of knowledge to young children:

> *The teacher displays sixteen sticks or counters and gives Emily the following instructions. "Make a group of five sticks. Are there enough left over to make another group of five?" (Yes) "Make as many groups of five as you can. How many groups of five are there?" (Three) "How many sticks are left over?" (One) "Are there enough sticks left over to make another group of five?" (No) "Okay, now group these counters by sixes. How many sixes are there?"....In this example, the teacher is giving a lesson on grouping objects into sets and naming the sets. This is thought to prepare the child for numeration and base representation concepts. The exercises require the kinds of exchanges that would be*

involved in operations in base 5 and base 6 (p. 105).

Secondly, changes in elementary school mathematics occurred due to the emergence of new theories about how children acquire mathematical concepts and understandings.

One of the most influential theories which shaped elementary school mathematics teaching was put forward by Dienes (1970). Dienes accepted the view that the acquisition of mathematical concepts is not simply a question of maturation and he attached great importance to the influence of the environment which, according to his view, must be carefully structured in order to help learning.

> *He strongly advocates the use of structured play situations in which the child moves gradually from a possibly intuitive grasp of the 'rules' (which Dienes does not despise, because he believes that much valuable learning begins this way) to a point at which the rules can be analysed and even manipulated to extend mathematical understanding.* (Taylor, 1976, p. 26)

Dienes distinguished six stages in learning mathematics. These stages are:

(1) The "Free play" stage - The child is provided with carefully designed games, but he uses them freely.
(2) Imposing constraints - The child realizes that games have rules, and that he has to obey these rules.
(3) Identifying underlying structures - The child plays various games and recognizes common elements among them.
(4) Representing abstractions - The child can not use abstractions and therefore he needs to represent them in some form of visual (or auditory) way, such as a graph or diagram.
(5) Understanding properties of abstraction - The child examines and understands properties of abstraction and describes them in his own language. It may be valuable for a child to invent his own language for the purpose of describing his own ideas.
(6) Devising formal systems - The manupulation of such systems is the final goal of mathematics.

Dienes claims that in "traditional teaching" children were taught the wrong way around. They were presented at a very early

stage of their studies with symbols, and only later were they shown through concrete materials, how to grasp the meaning of these symbols. The six stages of Dienes represent stages of acquiring a particular mathematic rule or concept, and whenever a child faces the task of acquiring new concepts, he must go through all these phases. It may well be that a child may reach the sixth phase with one concept, and at the same time need operations which apply to earlier phases for other concepts.

The idea of teaching mathematical structures rather than collections of computational rules has been commonly adopted in most innovative programs developed across the world since the 1950's. Nevertheless, the actual programs differ greatly from each other. The differences reflect disagreement over the best ways one may help children acquire mathematical structures.

Howson et al (1981) list several theoretical approaches to determining the sequence of instructional activities, and the organizational principles behind study materials and guidelines. To a large extent these approaches represent views about how the child does learn mathematics, rather than what he/she should know.

Among these the most importants are:

- The structuralist approach developed by Bruner, which is based on theories about concept formation.

- The behaviorist approach, which is based on learning theories. Programs which adopted the approach of structuring the learning processes around clearly defined behavioral objectives (Bloom et al, 1956), as well as those which adopted Gagne et al's (1962) approach of learning hierarchies pertain to this category.

- The formative approach - According to Howson et al (1981) this approach is formulated without reference to particular school subjects. "*It proceeds on two assumptions: first, that any school education aims at endowing the pupil with an optimal basic body of cognitive abilities and affective motivational attitudes; second that these factors may be described in terms of personality traits*" (p. 116).

- The integrated teaching approach - This approach

> is based on the same cognitive-theoretical basis as the formative one, however it considers also problems of content. The selection of content, according to this approach, should consider the needs and interests of the pupil.

It can be seen that the approaches listed by Howson et al(1981) are rooted in theories about learning and about the role of the school in society, rather than in conceptions about the nature of mathematics.

Resnick and Ford (1981) also employ a classification scheme based on psychological theories for classifying approaches to mathematics curricula. They describe curricula as being influenced by Gestalt principles, by theories of developmental-cognitive psychology, or, alternatively, by information-processing psychology.

Examining the classification schemes which attempt to characterize differences between mathematics programs in the 1980's, one may find categories derived from various branches of psychology. In reality, however, very few mathematical curricula can be fully described on the basis of a single guiding principle. Even in cases where authors explicitly specified the rationale of their work, one may identify elements in the instructional materials they produced by them, which are there mainly for pragmatic reasons, and are not related to, or justified, by the guiding principles they claim to adhere to.

It is of interest to note that the principle of constructionism, which in the middle of the 1980's is so dominant in describing principles of mathematics instruction, does not appear in the indices of the two leading publications on mathematics curricula (The books of Resnick and Ford and Howson et al published in 1981).

The concept of constructionism is defined by Pellerey (1985) as follows:

> *Mathematics is something referring more to a building up process where concepts, definitions, theorems, and their demonstrations are being fittingly heaped up, rather than to a logical structure pre-existing an understanding of it. There are no mathematical objects independent of the process of producing them. Therefore the*

> *goals in teaching are based more on promoting the ability to build up concepts, models, theories, and applications in mathematics rather than on reproducing already existing organizations. Mathematics is therefore someting more internal, intimately linked to ways of thinking, rather than something external.* (p. 3248)

And parallel to these theoretics refinements of epistemological, sociological and psychological perspectives determining the parameters of mathematics curricula, it can be also observed a movement of opposition to the idea of new or modern mathematics.

Vinner (1985) notes that the opposition to the new mathematics ideas reached its height in the middle of the 1960's. "*It was claimed that teachers and parents did not have the mathematical and pedagogical background to deal with it*" (p. 3511). He also indicates that educators complained "*...that children quite likely know that 3+5 equals 5+3, but they do not know how much 3+5 is*" (p. 3511).

Dissatisfaction with the outcome of mathematics education was expressed in the 1970's for various reasons. Berk (1985) refers to society's disappointment due to declining achievement test scores, as reported by national surveys of educational achievement (College Entrance Examination Board, 1977). Charges were raised against the child-centered approach in education, and against the pedagogy of permissiveness practiced in schools. But charges were raised also against innovative programs in general and innovative programs in mathematics in particular.

The disappointment with innovative mathematics programs was reinforced by educators, who were formally promoters of the innovative curriculum movement in the U.S. In his criticism of the new mathematics programs Sarason (1971) wrote:

> *...one of the bothersome features of the old curriculum was that children did not enjoy numbers; they found it dull, unstimulating, and a chore. To use more modern parlance: rather than being turned on, they were turned off, from mathematics.... observers [of classes where new maths programs were used] came away with the impression that enjoyment was one of the last words they would use to characterize their impressions of the feeling of children,* (p. 45)

The criticism of the innovative mathematics programs coincided with various manifestations of social crisis such as the weakening of the economy in the early 1970's, increasing violence in the schools, absenteeism etc., and created a cultural climate of nostalgia (Berk, 1985). In the wake of these feelings, schools began to appear which offered programs of limited scope --reading, writing, arithmetic, American history and geography-- together with didactic instruction of morals and patriotism. In these schools teachers unapologetically restored the rote methods.

These changes in the pedagogical climate became known as the "Back to Basics" movement in the U.S., and gradually it swept through most educational systems in the western world (Van Bruggen, 1985), and eventually reaching Israel, as well (Lewy, 1986). One of the pedagogical consequences of this movement was the introduction of Minimum Competency Testing in most American states and in several other countries (Jaeger and Tittle, 1980).

In elementary school mathemathics, the Minimum Competency Testing led toward greater a emphasis on the performance of rudimentary tasks in the domain of basic numeracy.

A historical perspective of mathematics teaching over the last hundred years is provided by Fey (1981). He pointed out two significant events: (a) the increasing importance of mathematics education around the world; and, (b) massive efforts to adjust mathematics curricula to the needs of industrialized societies and rapidly developing sciences, and actitivies of planning, designing, and implementing new or modified approaches of teaching and learning mathematics. He attributed these developments to the rapid "*industrial development, and the growing impact of modern technology on daily life, ... a growing political willingness to educate the masses, rather than just the elite*" (p. 2). However, he also noted that:

> *With all of this exciting change has come the less exciting observation that many children, even after attending classes regularly, have nevertheless not learned much of what the schools were expected to teach* (p. 3).

In an attempt to summarize the state of mathematics teaching, Davis (1985) addressed three broad problem areas of primary school mathematics education. He related these problems to the

"surprising" difficulty in answering the question "What should actually be taught in schools?". Firstly, he pointed out that there exists a great variation in mathematics curricula between schools, local districts, and nations. These national variations reflect each country's past and present socio-political and economical characteristics, conditions and needs --all of which define both the national policy of education in general, and, the policy of mathematics education in particular. Second, he noted that there are frequently discrepancies between what is "officially" described, expected and supposed to happen and what actually does happen. These differences are functions of diverse cultural characteristics, which determine the types of arithmetic teaching aids that are used (e.g., concrete materials or only symbols); the kinds of numerical symbol system(s) adopted; the adequacy of local languages or personal modes of language used for discussing issues in mathematics; and the adequacy of local and personal points of views for efficiently absorbing mathematical modes of thought, ideas and practices. Third, he observed that for answering the fundamental question previously referred to, it is mostly decisive the clear and meaningul identification of both the modes of teaching the mathematics subject matter, and, the ways the subject is learned.

However, it seems quite reasonable to assume that the diverse patterns of interaction existing within and between the three problem areas indicated by Davis (1985) engender the three basic ills of primary school mathematics programs, i.e., the relatively poor quality of these mathematics programs; the tendency to limit primary school mathematics education to attainment of a restricted set of cognitive goals; and the simplistic assumption sustained in a great majority of countries, that learning mathematics at the primary school level should requires memorizing facts and algorithms by rote, in contrast to both the "abstract mathematical structure" school of thought, and the "constructivist" position. These three weaknesses or flaws in primary school mathematics education are not isolated events, but are rather intimately related to each other. In more specific terms, they are flaws which function both as cause and effect in relationship to each other. They operate in a recurrent mode, each of them influencing and modelling the others dynamically. Furthermore, these recurrent interactions lead toward poor conceptions about <u>what should be the forms of teaching primary school mathematics, and how the process of learning mathematics actually occurs</u>.

In the 1980's two conflicting trends can be observed in

mathematics curricula in elementary schools. One strives to refine innovative mathematics programs and eliminate their flaws. This approach is strongly advocated by Resnick and Resnick (1980) who casted doubt on the efficacy of reviving rote methods of teaching. The other one strives to spread the idea of Minimum Competency Testing.

In Israel, too, one may feel the impact of these competing trends in mathematics teaching, though textbook writers and formally-approved mathematics programs are in favor of the ideas which underly the epistemological and psychological basis of the innovative mathematics programs. The second trend, i.e. the demand for core curriculum and minimum level of knowledge is reflected in discourses about mathematics teaching in in-service training programs, professional discussions, etc. It is a kind of oral lore, which has an increasing impact on teaching practice. The findings reported in the Van Leer study (Minkovits et al, 1977) gave an impetus to teaching computational skills in elementary schools. The annual Minimum Competency Testing introduced by the Ministry of Education in grade 3 of the elementary school (Razel et al, 1984) also reinforced this trend.

The present study, which deals with the assessment of mathematics achievement in grade four of the Israeli elementary school and was carrried out in 1985, bears signs of the two conflicting trends prevalent in the country. On one hand, it had to take into consideration the requirements of the formal curriculum, and the contents covered in the currently-used textbooks, and on the other, it had to satisfy demands for information concerning the mastery of basic computational skills in various substrata of the target population.

To a great extent, these foci of the study are referred to by Davis (1985), who contended that competing views will continue over the importance of "rote learning", "abstract mathematical structure", and "constructivist" views in the context of mathematics learning in the primary school. And the eventual adoption and implementation of any of them will inevitably shape mathematics study in primary schools for the next decade or so.

Additionally, the present study is justified, in that it attempts to increase the quality of educational opportunities for all learners. The increasing democratization of the world's population ought to have an impact on mathematics school curricula and their improvement, as stated by Davis (1985):

> *...if we are no longer to educate different classes or different castes, if the distinction between "gentleman" and "peasant" is to be eliminated (including also sex-typing as an important special case), what, then, is to be taught to every person? And what use shall be made of differences in individual ability? The answers to such questions remain to be worked out in the years ahead.* (p. 3244).

The Concept of Minimum Competency

The idea of "Minimum Competency" as a factor in shaping school curricula is by no means an innovation of our age. Schools for children of both peasants and the industrial proletariat have never striven to impart more knowledge than what may be defined as minimum competency. For a very long time, and in several countries till the middle of the 20th century, the whole course of schooling for these children was restricted to a few years, and approximately at the age of 10-12 children left school and entered the labor force. During these few years of school studies, children were supposed to learn the rudiments of reading and writing and of arithmetical computation. They also had to acquire some basic knowledge of the religious catechism, familiarity with national symbols, and frequently had to learn some labor skills. The utility of this knowledge justified its being taught in schools. It was believed that such knowledge would ensure that the recipient became a faithful member of the church and an obedient servant of the prevailing social order.

But even in societies where intellectual development of the child has been highly encouraged, and where opportunity to excell in scholarly studies has also been provided for disadvantaged children, it was assumed that those who have difficulties with the complexity of advanced studies should be taught less. Teachers ought to be satisfied if such children have acquired some rudimentary knowledge of basic skills.

Lewy (1985b) mentioned the common-sense pedagogical principles of the Jewish school-teacher described in a novel by the Hebrew author Y. Burla:

> *The hero of the novel Aqavyah, would be defined in*

> *contemporary terms as a disadvantaged learner. When he was taught to deliver his Bar-Mitzvah speech the teacher designed for him a question and answer type oration that could be memorized easier than the customary monologue. The innovative feature of these programs is not their pegagogical message, however, but rather the fact of rendering to it statutory power* (p. 315).

The emergence of the Minimum Competency concept in the 1970's as a powerful principle for determining school curricula has been extensively discussed in the professional literature. A through description of a broad variety of approaches to Minimum Competency Programs is provided by Gorth and Perkins (1979).

Approaches to Defining Minimum Competencies.

The common feature of these programs is that schools should assume responsibility for imparting basic knowledge for all children who enroll, and that schools should issue graduation certificates or diplomas only to those who have proved that they mastered a satisfactory level of knowledge of what has been defined as Minimum Competency requirement.

As to the content of the Minimum Competency requirement Gorth and Perkins (1979) described three distinct approaches: basic skills, life skills and syllabus-oriented approaches.

The Basic Skills Approach - The basic skills approach emphasizes mastery of basic literacy and numeracy, and places emphasis on acquisition of the traditional three Rs: reading, writing and arithmetic. This approach received great support from the Back to Basics movement in the United States (Brodinsky, 1977). This movement, though it initially received no endorsement from any organization, received strong support from the Council for Basic Education, which began operation in 1975. In the wake of this movement, we witnessed a renewed debate between the supporters of the progressive and traditional, or formal, educational philosophies.

The Life Skills Approach - The Life Skills approach to defining Minimum Competencies emphasizes the utilitarian aspects of school learning. Schools have to prepare the learners for successful functioning in our increasingly complex society. Riehs (1985) indicates that there has been a tendency among curriculum

planners to differenciate between three categories of core competencies: basic skills (including literacy and numeracy which are preconditions for further learning), life skills, and role-preparation skills. The survival skills touch upon areas of consumer behavior, parenting, interpersonal relationship, mechanical skills, financial orientation, work skills, leisure skills, and citizenship skills.

Syllabus Oriented Approach - A third common approach to minimum competencies is based on the analysis of the curriculum, or documents (syllabuses) which specify the curricular objectives. Minimum competencies are defined by scrutinizing curricular objectives listed in the syllabuses, and by distinguishing between these requirements and requirements beyond these boundaries. This conception of Minimum Competency is based on the assumption that within each society a satisfactory level of consensus can be obtained in regard to the core elements contained in each syllabus. In other words, schools in a particular society which cannot cover all curricular objectives contained in a syllabus, and which have to omit certain objectives from their actual program, will nevertheless agree on what are the core elements, and will teach these elements even within the framework of a reduced program. This does not mean that an educational system should rely on the intuitive judgments of policy makers at the school level. On the contrary, a syllabus-related Minimum Competency Program has to specify the core elements and has to ensure that all schools adequately teach them. Still, there is an assumption that the specification of the core elements is not an arbitrary decision made by policy makers, but rather a guideline for content selection, which enlightened teachers would make even without formally specifying a set of recommendations.

The Emergence of Minimum Competency Programs in the 1970s

Minimum Competency Programs were introduced to schools in the United States and other industrialized countries in the middle of the 1970's. Larkins (1981) indicated that the term Minimum Competency is fully absent from the educational literature in 1970. The December 1970 issue of the Phi Delta Kappan, devoted to the topic of accountability, did not even mention the term Minimum Competency. The 1970's was the decade of legisilation in the U.S. on matters of Minimum Competency testing, and it reached its peak towards the end of the decade. In 1980, 42 states (of 50) had legal requirements for some kind of Minimum Competency testing

(Baker and Choppin, 1985). Tittle (1982) studied the reasons which led to the emergence of Competency Programs in the U.S. She identified three major factors: the testing context, the economic and the accumulated evidence revealing deficiencies in educational achievements nationwide.

The testing context - Educational systems established testing services a long time ago. The Regent Examinations in New York were established for elementary schools as early as 1865. But developments in educational psychology, test theory, and measurement technology in the second half of our century were especially conducive to an increase of tests utilization in schools as a device of controlling curriculum. The idea of Minimum Competency became influential when increased attention was paid to issues of defining educational objectives (Bloom et al, 1956). Glaser (1963) reinforced the view that a strong relationship should be established between educational objectives and instructional practice by laying the foundations for Criterion Referenced approach to testing. The efforts made in the U.S. to promote equality in the field of education also contributed to the system's willingness to pay greater attention to the needs of those who, under regular conditions, failed in the schools. Finally, development in the technology of Mastery Learning (Block, 1971) gave a boost to the introduction of Minimum Competency Programs and Minimum Competency Testing. The Mastery Learning technology proposed that under appropriate learning conditions, which include the utilization of frequent and detailed formative feedback about attainment of specified instructional objectives and the provision of immediate corrective teaching, it is possible to help almost all children in school to successfully master prescribed school objectives. Indeed Baker and Choppin (1985) describe the Minimum Competency Programs in the U.S. as large-scale or system-wide Mastery Learning projects.

The economic and social context - in the United States the rapid growth of the economy in the 1950's and 1960's resulted in the establishment of social programs which aimed at what Cohen and Haney (1980) called "minimum goals" in areas such as Public Housing, Social Security, and Medical Care. The idea of egalitarianism was fostered in the field of education as well. Secondary education, which was previously an elitist educational system, expanded and opened its gates practically to all pupils of any age group. In most industrialized countries there are no admission requirements to secondary schools. Consequently, the curricula also had to be adjusted to the needs of non-academically oriented

pupils.

National deficiencies in educational achievements - The third factor which precipitated the establishment of Minimum Competency Programs was the evidence garnered from research studies on the measable decline in scholastic achievements. Both national and cross-national assessment studies produced such results. One of the first signs of the alleged deficiencies in the American educational system appeared in the results of a large scale and multiphase crossnational comparative study carried out by the International Association for the Evaluation of Educational Achievement (IEA), a consortium of leading research institutes from approximately 40 countries. The first study dealt with achievement in mathematics and its results were published in 1967 (Husen, 1967). Israel participated in this study Israel, also, and the Israeli national average was among the highest level averages in the study. Nevertheless, doubts were expressed about the adequacy of sampling procedures in Israel, and Israeli researchers received the the results touching upon the obtained achievement in Israel with scepticism.

In contrast, the American achievements were very low. Newspapers strongly reacted to this finding. Husen (1977) wrote that in the U.S.

> *...more than one hundred press chippings bear witness to the concern caused in the United States by the standing of that country in the "international mathemathics league"....Both Congress men and newspaper editorials in the United States raised the question whether all the money spent on the overhaul of mathematics curriculum and on the research on the proper way teaching school mathematics has been wasted. Persistent critics of American education saw the low mean performance as an indication of the low quality of American education in general.* (p. 132)

From within the United States, more evidence about the decline of achievement scores came from several different sources. We will mention here two of them: The SAT scores and the NAEP results.

The Scholastic Aptitude Test (known by the acronym SAT) is a screening test administered to high school graduates who wish to apply for admission to colleges and universities. The test is

administered year after year to new cohorts of high school graduates. The summary of the test scores exhibited a trend of decreasing averages. The 1970's were characterized by a continuous and considerable decline. There are reports of a continued decline in 1980, though its magnitude is relatively low (College Entrance Examination Board, 1980).

The NAEP (acronym of the National Assessment of Educational Progress) is an institute established by the Educational Commission of the States, an official coalition established as a means for the states to work together on common educational interests. The aim of NAEP is to periodically provide information about the educational achievements of the younger population of U.S. (Tyler, 1985). The NAEP carries out assessment studies through of national samples of pupils at the ages of 9, 13 and 17. It also carries out assessment of young adults in the age range of 26-35 years.

The results of NAEP indicate that during the 1970's declines occured for all age groups in science, and in mathemathics for 17 year-olds. No changes were reported between 1971-75 for the 13 year-olds and 17-year-olds, and a slight improvement was recorded for the 9 year-olds in reading skills. There was slight improvement in literal comprehension and a slight decline in inferential comprehension for the 13 year-olds and 17 year-olds (Farr and Olshavsky, 1980).

Criticism of the Minimum Competency Testing

In the United States severe criticism was uttered against the practice of Minimum Competency Testing. One of the sharpest criticism appears in Larkins's (1981) work. He considers MCT as a fad which is doomed to have a short life. Larkins observed several symptoms of faddism in the Minimum Competency Testing Programs. First, they exhibit the characteristics of the "bandwagon". Schools, districts, states moved quickly and uncritically to climb on board. Secondly, publications about the topic have unscholarly characteristics. They lack the marks of scholarship such as footnotes and bibliographies. The authors rarely seem aware of the historical perspective of the concept. Finally, there is a lack of evidence. There are no empirical studies which validate the effectiveness of these programs. But claims are made that programs of this kind will improve educational achievement, that school diplomas will acquire

meaning, and that morale among teachers, students and parents will increase.

Haney and Madaus (1978) indicate that there is no real evidence of the reported achievement test decline. Decline in the SAT scores may be attributed to changes in the socio-economic composition of those who apply for admission to colleges. Additionally, the increasing diversification of secondary school mathematics curricula may have created a substantial difference between what is taught in schools and what is examined by the test. In face of this curricular diversification, it may now be impossible to develop a test containing items which deal with topics that most of the students in the target population had an opportunity to learn in the schools.

Moreover, the NAEP scores indicate that the decline of achievements, if indeed there is a decline, is not caused by a decrease in minimum competencies, but is due to the diminishing knowledge of advanced level. As already indicated, there was a slight improvement in the field of literacy comprehension of texts and a slight decline in inferential comprehension. Thus, it may well be that the Minimum Competency Programs lead to an increase in teaching in areas where no decline has occured. Cohen (1976) believes that the claim of declining scores is no more than nostalgia for times past, or a sense of loss.

In the face of such critical comments on the Minimum Competency Programs in the United States, it seems necessary to have a closer look at the developments in the field of Minimum Competency during the last decade.

Approaches to Minimum Competency in Israel

Since the present study was heavily biased toward an examination of the mastery of minimum competencies in fourth grade mathematics in Israel, an examination of the relevant approaches in the Israeli educational system seems appropriate.

As indicated above, the educational literature which treats the subject of Minimum Competency does not pay attention to the historical perspectives of such programs. According to Lewy (1985b) the Israeli educational system has twice faced the challenge of introducing some kind of Minimum Requirement Programs. The first was in the 1960's, i.e. long before the

Minimum Competency Testing movement started in the U.S. In any event, when in the 1980's Israel jumped on the bandwagon and decided to introduce Minimum Competency Programs, it was considered to be an innovative attempt and no reference was made to its predecessor.

The 1960's program came to be known by the name "Parenthesis Program" due to the fact that the Ministry of Education and Culture produced a reduced program in various subject areas by putting into parenthesis (and thus eliminating from the list of obligatory study) those program elements which seemed to have only secondary importance. The 1960's MRPs were neither implemented nor granted statutory power, nevertheless they had a decisive impact both on defining learning objectives, and on devising learning activities for those learners who were considered to be "educationally deprived", and for those who were considered as being able to answer all the demands of the regular school program as well (Lewy, 1981, 1985b).

Indeed, the 1980s' Minimum Competency Program reflects a number of characteristics which qualified the 1960's program. Both attempts tried to accomplish the same aim of "*...improving the achievement level of educationally disadvantaged children*" (Lewy, 1984b,p. 1). Also, in both attempts "*...the program goal was the same: to ensure that even the 'weakest' groups of the learner population acquire a level of basic knowledge, judged as being essential in order to function adequately in a modern society*" (Lewy, 1984b, p. 1). Additionaly, both them are characterized by Lewy (1984b) as being based on a "backward look" towards existing programs.

For the sake of comparison an excerpt (translated) from the 1960's and the 1980's Minimum Competency Program in mathematics are presented is Exhibits 1.4 and 1.5.

It can be seen that in the 1960's the concern of educators was to reduce the computational burden imposed upon "weak" students, while in the 1980's the concern was to eliminate complex conceptual topics or to slow down the pace of content coverage for "weak" learners. In the 1980's program, priority assignment is given to topics and it is recommended that priority B and C topics should be taught only in classes in which students have already mastered priority topics marked A.

The concept of Minimum Competencies became influential in

Grade Four Syllabus

Topic	No. of hours of teaching	Priority level
The decimal system	10	A
Multiplying by 10	2	A
Signs of divisibility by 2, 5, 10	3	B
Review of basic multiplication	10	A
Multiplications of columns	6	B
The order of operations and parenthesis	3	A
Rounding numbers for estimating quantities	4	B
Fractions (oral and symbolic representation)	4	C

Exhibit 1.4 Excerpts from the grade four syllabus 1982

Excerpts* from the 1957 Arithmetic Syllabus Grade Four

Comprehension of map scales [mainly] in a passive way.

Familiarity with the Hebrew [and the regular] calendar.

Reading and writing integers up to [one million] 100,000.

Addition and Subtraction of writen numbers up to [one million] 100,000.

Finding the 1/2, 1/4, [3/4], 1/3, [2/3], 1/10 of quantities and of small numbers which are exactly divisible.

*The excerpts contain items in which parenthesis is present

Exhibit 1.5 Selected items from the parenthesis program

achievement testing in Israel, too. The office of the Chief Scientist of the Ministry of Education carried out Minimum Competency Testing in grade 3 of the Israeli schools. The purpose of these testing programs was to identify those pupils who have not mastered basic reading and computational skills and to prescribe remedial treatment for them (Razel et al, 1984). At the same time, the Curriculum Center of the Israeli Ministry of Education launched a series of assessment studies in various subjects, and embedded a Minimum Competency Test in a general subject matter test which was designed for collecting information about the state of the art of achievements in a particular subject in the school system. Till now, such assessment studies were carried out in Science (Lewy, Bacher, and Raz, 1983), in Biblical studies (Padua, 1986), in History (Wolff, 1983), and in Geography (Raz, 1986).

The assessment study in mathematics reported in this work constitutes a chain in the series of these studies. The test used in this study, too, has a set of 10 embedded items examining the mastery of Minimum Competencies.

The specific features of the test, and the embedded series of Minimum Competency items, will be described in the Chapter on data-collecting instruments.

Chapter Two
The Operational Framework

The Goals of the Study

The primary goal of the study is to provide information about the state of the art in Israel of teaching mathematics at the elementary school level. A second major goal of the study is to explore the feasibility of establishing and maintaining a multipurpose feedback system, and to find out in what ways the effectiveness of such a feedback system could be ensured.

In more general terms, it can be said that the study's contribution touches first upon substantive aspects of mathematics education in the elementary school, and secondly on management and methodological aspects related to multipurpose feedback systems. Accordingly, the two aspects of the study, that which addresses substantive aspects of mathematics teaching, and that which addresses management and methodological aspects of feedback systems, will be dealt with interactively in order to further the attainment of the study's two major goals.

Substantive Aspects

In the domain of the substantive aspects of mathematics instruction, the following specific issues will be addressed:

(1) Intended mathematics curriculum - A portrayal of the intended mathematics curriculum will be presented on the basis of a critical analysis of the national syllabus and of instructional materials such as textbooks, complementary and supplementary instructional aids.

The analysis will focus on basic principles, aims, and instructional strategies and planned activities, as they are

represented in the previously mentioned documents.

(2) Implemented mathematics curriculum - The implemented curriculum will be examined in terms of two major clusters of variables which are referred to by Fullan (1981) as (1) "conditions of implementation" variables, and (2) teachers' and learners' program-related activities. The first cluster of variables will contain information about curriculum inputs such as: (a) pupils' socioeconomic origin; (b) teachers' professional background; and (c) time related variables.

The second cluster of variables will deal with activities which are directly related to using instructional materials in the class. Special attention will be paid to the Opportunity to Learn variable (OTL), which has been systematically used in various IEA studies (Husen, 1967; Anderson, 1985).

The results derived from the collected data will be summarized in a way which will enable us to portray the following aspects of the implemented curriculum:

(2a) Distribution patterns of selected "conditions of implementation" variables;

(2b) Distribution patterns of "actual teaching-learning process" variables;

(2c) The pattern of relationships between the Opportunity to Learn variable and the intended curriculum.

(3) Attained mathematics curriculum - Both cognitive and affective outcome variables will be dealt with.

Cognitive variables will be portrayed in terms of pupils' attainments in the areas of routine computation, mathematics comprehension, mathematics problem-solving, as well as in terms of pupils' mastery of minimum competencies in mathematics.

Affective variables will focus on the following types of pupils' attitudes: liking mathematics; interest in mathematics; perceived importance of mathematics; difficulties encountered in studying mathematics; mathematics anxiety.

The results derived from the collected data will be dealt with so as to answer the following questions:

(3a) Which instructional goals specified in the intended curriculum were attained?

(3b) What differences can be observed between schools and among individuals within schools with regard to the attainment of various instructional goals in mathematics? More specifically, information will be provided about achievement differences across administrative districts, gender of students, and users of various educational programs.

Methodological Aspects

(4) Interdependence of assessment and program evaluation - The study will strive to offer evidence that the traditional conceptions about differences between assessment and evaluation studies are not necessarily valid, and that there are ways of reducing the conceived dichotomy between assessment and evaluation practices.

This means that one can derive information from a single set of data to fulfill roles, which, in the past, have usually been distinctly attributed either to assessment or to evaluation studies.

To attain this goal, an attempt will be made to demonstrate the actual utilization of summaries derived from a single set of quantitative data for assessment and for program evaluation (and therefore for decisions) both at the whole system and at the classroom levels.

As already indicated, the data summaries will serve as a basis for formulating recommendations on how to improve mathematics programs currently employed in the Israeli primary school, as well as mathematics teaching and learning modes and processes.

The Management of the Study

The study focusing on the state of the art of mathemathics teaching in Israeli elementary schools was initiated and carried out by the Israel Curriculum Center of the Ministry of Education and Culture. According to the regulations of the above-mentioned body a Steering Committee was established to monitor the study and to supervise activities at various phases of the study. The

members of the Steering Committee were nominated by the head of the Israel Curriculum Center (ICC). The Chief Superintendent for Mathematics Teaching of the Ministry of Education, the director of the Mathematics Curriculum Development Team at the ICC, representatives of other mathematics curriculum projects in Israel, and a representative of fourth grade elementary school mathematics teachers participated in the Steering Committee and worked out the reseach plan for the study, and monitored the development of the instruments, the data collection and the data analysis. They also participated in decisions related to the dissemination of the results to various "stakeholders" of mathemathics teaching in Israel.

The writer of this report participated in the whole process of planning the study and carrying it out. The data collected within the framework of the study were summarized by the Evaluation Team of the Israel Curriculum Center, according to the design which fitted their interest. The data collected by the Evaluation Team served as a data base for a secondary analysis carried out by the writer. The pattern of data analysis presented in this report was determined by the writer independently and without any relationship to the work carried out by the Evaluation Team, and without being monitored or supervised by the Steering Committee.

The Population and the Sample

The Population

The population of the study is defined as the total learner population in the fourth grade of the primary school in the Israeli Hebrew educational system, excluding those children placed in special education classes.

The decision to carry out the evaluation study at the fourth grade level was made on the basis of certain priorities determined by the Ministry of Educational and Culture, and supported by the Chief Inspector for mathematics instruction in the Israeli educational system, and by the Head of the Mathematics Team in the Israel Curriculum Center. The Ministry of Education and Culture was curious to know how well the newly-developed mathematics programs for the lower grades of primary school fit the needs of the target population. Since most of the newly-developed mathematics programs in Israel were gradually implemented in the system,

starting with the first grade and moving forward year after year, it is only now (after four years) that the great majority of these programs have been completed for grades one to four. Moreover, the fourth grade level seems to be a stage of mathematics learning wherein the most fundamental concepts and skills should have already been acquired.

The Sample

The sampling unit was the school. Schools were selected by the method of stratified random sampling. In previous assessment studies carried out in Israel, the stratifying variables were the six school districts operating in the country; the compensatory treatment index of the school; and the size of the school (Lewy and Chen, 1977; Wolff, 1983). To these commonly-used stratifying variables a fourth variable was added in the present study: the mathematics instructional program (textbook and/or teaching aids of other type) used in the school. Information is available on the first three stratifying variables in the administrative records of the Ministry of Education. No centrally kept record is available about the textbooks used in schools. Therefore it was necessary to carry out a two-stage sampling plan. In the first phase, information was gathered about the textbooks used in school and this information served as an additional stratifying variable for drawing the final sample. The size for the first sample was 300 schools. This sample size provided, at a relatively low cost, information about instructional programs used in schools. The second phase involved intensive data collection in schools and therefore the sample size was reduced to 100.

The reasons for selecting the above-mentioned four stratifying variables and the process of the two phases of sampling will be described in the subsequent section.

The Stratifying Variables

Administrative districts - From the organizational point of view, the educational system in Israel is divided into six geographical districts. Recently, the Ministry has striven to strenghten district autonomy. Consequently, separated information about the operation of the schools in each district was desirable.

The department of planning of the Ministry of Education

publishes basic statistics about schools each year, and in this publication the schools are listed according to school district. Using this variable for the sake of stratification was therefore both a matter of convenience for those who draw the sample, and useful for the recipients of the information to be derived from the study.

The compensatory treatment index - The Ministry of Education provides funds to finance compensatory educational programs within schools. Schools are entitled to benefit from these funds to the extent that they provide education for children belonging to educationally deprived social groups. The Department of Planning annually rates the schools on a scale ranging from 0 to 100, based on the proportion of children from educationally deprived groups studying in that particular school.

This index, which was developed for the purpose of allocating compensatory education funds, serves as the best indicator of the average ability and achievement level of pupils studying in it. In order to get accurate estimates of national parameters it was useful to stratify the population according to the compensatory treatment index assigned to schools.

The size of the school - The third stratification variable used in numerous assessment studies in Israel is the size of the school. There are two reasons for employing this variable for the purpose of stratification. First, there are good reasons to believe that the size of the school makes a difference from the point of view of the availability of resources. In Israel, as well as in many other countries, a large school has at its disposal better equipment and richer human resources than small schools. The second reason is statistical in its nature. As demonstrated in numerous studies (Lewy and Chen, 1977; Lewy et al, 1978) Israeli schools are highly internally homogeneous, i.e. the variation in achievements across schools is relatively high in comparison to variations inside the school.

Therefore, keeping constant the number of students in the sample, it seemed useful to minimize the number of pupils sampled within one school, and to maximize the number of schools included in the sample. Following Peaker's (1967) advice, it was found that a sample of 30 persons from a single school provides a quite stable estimate of the parameters of a single school's population.

Moreover, Peaker (1967) proved that in an educational system

like Israel, in which there is a relatively high level of homogeneity within elementary schools (approximately .15 intraclass correlation) increasing the number of persons sampled in a single school contributes relatively little to the precision of the estimates. He concluded that adding a second class to the sample from a school which is already represented would contribute less than 50 percent accuracy to the precision of the statistics than drawing one class from another school. In very rough terms, this means that the precision of estimates obtained from 100 classes taken from 50 schools, equals the precision obtained from 70 classes, each one coming from a different school (see Peaker, 1967, vol. I, p. 149).

Therefore, it was decided not to sample more than 30 students from a single school. Thus, the resources were employed in increasing the number of schools contained in the sample, rather than in increasing the number of pupils in a single school.

In this respect the sampling plan was identical with that employed in a Israeli IEA study (Lewy et al, 1978). Nevertheless, in the IEA study a group of 30 students was randomly selected from all parallel grade levels operating in a single school. Since such an arrangement imposed difficulties on the data collection, and the routine work of the school was disturbed in all classes from which students were randomly selected, it was decided to sample a whole class in each school. The selection of the class within a particular school was done by the central research team, thus schools could not avoid that "weak" parallel classes should be excluded from the sample.

It turned out that in small schools where there was only one fourth-grade class all students were included in the sample, and in larger schools where more than one fourth grade class operated, only a fraction of the fourth grade was sampled.

Table 2.1 contains information on the distribution of schools according to their size.

In order to obtain a sample which provides a direct estimate of population parameters (without the need to assign differential weights to various strata), it was decided to use smaller sampling intervals for large schools than for small schools.

For practical reasons schools were divided into two groups. Small schools were considered to be those in which one or two

Table 2.1 Schools according to the number of parallel fourth grade classes in the whole Israeli School System.

Number of parallel classes	Number of schools	Number of classes	
1	447	447	806 schools, 1165 classes, average: 1.44 classes per school.
2	359	718	
3	193	579	331 schools, 1185 classes, average: 3.58 classes per school.
4	98	392	
5	29	145	
6	9	54	
7	1	7	
8	1	8	
All	1137	2350	

parallel fourth-grade classes operated, and large ones were those which had more than two parallel classes. As indicated in Table 2.1 small schools had an average of 1.44 classes and large schools had an average of 3.58 classes. To obtain a sample of 300 schools it was decided to utilize a sampling ratio of 4.66 for small schools and 2.59 for large schools[1].

The 1137 schools included in the target population appear in

[1] In practice two vectors of numbers were constructed. One list started with the number 4.66 and continued by 9.32, 23.98, 18.64, etc., i.e. always increased by 4.66. According to the same algorithm the other vector was 2.59, 5.18, 7.77, 10.36, et. The integers appearing in the vectors indicated the ordinal position of a school in an appropriate list, for being drawn into the sample.

the Records of the Ministry of Education in sequential order according to administrative district. For the purpose of sampling a new Record of Schools was produced in which schools were listed according to the following hierarchical nesting: size of schools, compensatory education budget index, administrative school district. This arrangement put into one group the small schools and into another group the large ones, and thus it only required changing the sampling interval once.

The actual sample obtained for the first phase is described in Table 2.2.

It can be seen that the sample distribution closely followed the distribution of all schools in the system. It should be noted however that the discrepancy observed between the distribution of schools in the sample and in the total population not only represents sampling error, but is also due to the fact that the sampling ratio was not the same for small schools and for large schools. Thus the number of schools appearing in the sample column represents the number and the percentage of learners studying in a school of a particular "compensatory budget index" range, and not schools pertaining to that group, while the percentages appearing in the column of total population refer to schools.

Out of the 300 schools which were asked to respond only 272 did

Table 2.2 A sample of 300 schools.

Compensatory Education budget index	Sample		Percentage in total population
	number	pct	
0 - 20	115	39	37
21 - 40	56	19	17
41 - 60	36	12	15
61 - 80	39	13	12
81 - 100	22	7	9
No information	32	10	10
All	300	100	100

so. These 272 schools constituted the pool from which a smaller sample of 100 schools was selected for further data collection on conditions of teaching mathematics and on achievements in mathematics.

In order to ensure that all textbooks used in the system should be appropriately represented in the sample of the 100 schools, a list was prepared in which the 272 schools of the first phase sample was rearranged and put in a new sequential order. In this new order the first classification variable was the textbook used in the school. Within the categories of the schools using a particular instructional program, the order of the schools was not changed. This ensured that the prior, nested classification of schools according to size, compensatory index, and administrative district, remained unchanged. To avoid an over-representation of small schools it was decided to consider the size of each school in the group of the 272 schools and to assign each school as many "tickets" as the number of its parallel grade four classes. Thus, a school which had only one fourth grade class received one "ticket", and that which had two fourth grade classes received two "tickets". Counting the tickets it turned out that the 272 schools had altogether 558 "tickets", or in other words they had 558 grade four classes. Since the intention was to sample 100 schools (and one class in each school) it was decided to use a sampling ratio of 5.58.

The distribution of the sample of 100 schools according to the compensatory index is presented in Table 2.3.

As indicated in Table 2.3, in the final sample a larger representation is given to low compensatory index schools.

The reason for this is quite likely the fact that among the 28 schools (out of 272 schools) who did not respond to the initial questionnaire there was an overrepresentation of high compensatory index schools. Nevertheless, it was decided not to change the sample. Due to the slight bias in the sample, consideration was given to the possibility of attaching differential weights to the results obtained in various intervals of the compensatory index and thus eliminate the impact of possible bias. But it turned out that the "compensatory index" had a relatively low correlation with test scores, and therefore any weighting of the results appeared to be unnecessary.

The final sample consisted of 100 schools. From one school the

Table 2.3 The sample of 100 schools

Compensatory index	Percentage of 272 schools	Percentage of 100 schools
0 - 20	39	47
21 - 40	19	23
41 - 60	12	11
61 - 80	13	11
81 - 100	7	3
No information	20	5
All	100	100

data was not returned to the Center, and therefore the study is based on data collected in 99 schools only.

The Data Collection Instruments

Three data collection instruments were used in this study: Teacher's Questionnaire (TQ), Pupil's Questionnaire (PQ), and the Mathematics Achievement Test (MAT).

Teacher's Questionnaire (TQ)

The teacher's questionnaire contained professional background-related questions and implementation-related questions. The professional background-related items (BKG), which are relevant to the teaching practice, were directly or indirectly related to the program implementation. The major clusters of the implementation-related questions are: in-service training activities related to teaching mathematics (INS); the teacher's autonomy in matters related to teaching mathematics (AUT); attitude toward the instructional program (ATT); initiatives of the teacher related to program implementation (IMP); and, opportunity to learn skills and topics included in the achievement test (OTL).

Pupil's Questionnaire (PQ)

The pupil's questionnaire contained a few biographical items (gender, and father's educational level). We did not want to pose questions about the country of origin of the pupil or of the parents, since this might have embarassed the pupil and/or teacher. To avoid such embarrassment, the information about the educational level of the pupil's father was recorded on the pupil's questionnaire by the teacher after the pupil completed it. The teacher obtained this information from the pupil's school file.

The questions posed to the pupils clustered around four issues: attitude toward mathematics; parent's interest in their child's study of mathematics; mathematics anxiety (phobia); and, access to, and utilization of pocket calculators.

Illustrative items from the pupil's mathematics attitude test are:

- Do you like studying mathematics?
- Would you like to enter a vocation in which mathematics is intensively used?
- If a mathematics lesson is cancelled how do you feel about it?

Items dealing with attitude in mathematics were clustered into a single scale. It contained 9 items and its reliability (coefficient alpha) was .720. (The serial numbers of the items contained in this scale are: 23, 29, 30, 31, 32, 33, 34, 35, 39).

The Mathematics Achievement Test (MAT)

Initial considerations

The final version of the mathematics achievement test contained 40 items. The items of the test were compiled by the evaluation team of the Israeli Curriculum Center (ICC) under the guidance and the supervision of the Mathematics Research Committee. The starting point in the compilation of items for the test was the mathematics objectives presented in the syllabus (Ministry of Education and Culture, 1972) and in the textbooks of various mathematics projects, and also a pool of test items available at the Israel Curriculum Center.

The syllabus lists, among others, the following objectives:

1. Acquisition of mathematical thinking, including the ability of inference, generalization, etc.
2. Ability to carry out mathematical operations with integers and fractions and understanding their meaning.
3. Knowing mathematical laws and being able to apply them.
4. Familiarity with geometric figures.
5. Comprehension of the mathematical aspects of daily life.

Items to be included in the test were suggested by the ICC's mathematics project team and analytically scrutinized by the evaluation team. The project team assumed responsibility for the content validity of the test and the evaluation team examined the items from the point of view of compliance to rules of item generation (such as clarity of the item stem; length of the distractors, etc.).

During the process of test development two major problems emerged, constituting a matter of controversy between the people involved in its construction: the balance between "strictly numerical" and "word" questions, and the matter of response mode. The pool from which the items were drawn had a high proportion of "strictly numerical" items and thus the early versions of the test had an axcessively high proportion of numerical items.Some members of the Mathematics Research Committee argued that the proportion of "word" problems should be increased, while others argued that by increasing the proportion of "word" problems one takes the risk that the test results will be biased due to differences in reading ability among the respondents. Finally, after gathering experts' opinion, it was decided to increase the number of word problems.

The second controversy touched on the mode of the item responses. The mathematics project team argued for open-ended questions, claiming that multiple-choice questions are not appropriate to test skills in mathematical computation and analysis. These arguments were refuted by other members of the Committee. In any event, the decision in favor of a multiplechoice response mode was made on the basis of technical considerations, rather than as a result of resolving the above-mentioned controversy.

The decisions made at the initial phase of planning the test demonstrate how complex is the process of determining basic features of a test when it is designed to serve a whole educational system. It also demonstrates, in such cases, the impact of organizational and financial considerations on decision making (McGrath, 1982).

Examining the quality of the test.

The preliminary version of the test was examined in two different ways: tryout in a small judgmental sample of the target population, and judgment by representatives of various mathematics project groups whose textbooks are used in the system. Later, the content validity of the final version of the 40 items math test was examined by using a two dimensional behavior-content grid.

(1) Tryout of the pilot version - The pilot version of the test was tried out in 4 classes. The classes for tryout were selected on judgmental basis in a way which yielded a representative sample of the target population.

The responses of approximately 120 pupils were submitted to item analysis and items which had a low discriminations index (lower than .250) were eliminated. No easy items were eliminated from the test since summarizing some selected set of items in a criterion-referenced manner was intentional. Therefore, the inclusion of very easy items in the test seemed to serve the purpose of the investigation.

(2) Judgments by the representatives of various mathematics projects used in the system - As a result of a preliminary exploration it was found that 22 different sets of instructional materials were used in the system. Four of these sets of materials had a relatively high diffusion in the system (at least 10 percent of schools used those materials). In order to avoid bias in favor of, or against some set of instructional materials or textbooks, the representatives of four major projects were asked to examine the test and to rate each item on a scale which specified the level of the treatment the item (or the content or skill represented by the item) obtained in their project. This three level scale had the choices: intensively treated, fairly treated, and not treated.

Additionally, the representatives of the projects were requested to identify topics and skills, if any, which were treated in their books, but were not represented in the test.

(3) Content validity examined through a two dimensional grid - The utilization of two dimensional grids for examining the content validity of tests as well as for planning instruction is highly recommended by Bloom et al (1971). Their book contains grids for a variety of school subjects. In these grids one dimension

represents instructional contents, and the other relevant mental behaviors. In contents, four areas were easily identified: number system, integers, fractions and geometry. In the behavioral dimension, a decision had to be made in favor of one of the commonly-used classification schemes in dealing with mathematics study in schools. It was decided to adopt the National Longitudinal Study of Mathematics Achievement (NLSMA) Scheme (Wilson, 1971). This decision was made due to the simplicity of the scheme (it contains only four categories) and its resemblance to the widespread taxonomy scheme of Bloom et al (1956).

We will present here the behavioral dimensions only (quoted via Pellerey, 1985):

(a) Computation: this level refers to recall of fundamental mathematical facts, terminology, and the skills required in order to carry out algorithms. The accent is placed on the simple evocation of knowledge and on fulfilment of operations in appropriate ways.

(b) Comprehension: this category covers remembering concepts and mathematical generalizations. The student must show comprehension of concepts and relations among them; translate data from one mode to another.

(c) Application: this involves the solution of familiar problems similar to those the student had to face during the process of learning. Pupils have to choose an algorithm and a solution and apply it succesfully.

(d) Analysis: the student must go beyond what was done during previous instruction, meeting new math experiences, solving uncommon problems.

The reader may recognize the traces of Bloom et al's (1956) Taxonomy of Educational Objectives. Indeed the NLSMA's categories constitute an adaptation of the general Taxonomy to the mathemathics discipline. The concept Computation is substituted for the original Knowledge category in Bloom et al's Taxonomy. Nevertheless its more detailed description is quite close to the Taxonomy's definition. "Recall of fundamental facts and terminology" is explicitly listed as one of Bloom et al's subcategories.

The other three terms, Comprehension, Application and Analysis are directly borrowed from Bloom et al's Taxonomy and the definitions, as well, are very close to those presented in his Taxonomy. One exception is the definition of Analysis. In the NLSMA list the term analysis comprises mental activities, which in the Bloom et al's scheme would be labelled as Synthesis. Quite likely, the

paucity of synthesis items which can be included in a mathematics test motivated the NLSMA expert to collapse this category with analysis.

The figures appearing in the various cells of Table 2.4 indicate an acceptable level of balance among the items (the ordinal number of the test items pertaining to the various behavior dimensions by content categories appear in Appendix I). According to the content categories the integers received the fullest coverage (26 items). Fractions, according to some projects, is a relatively new topic in grade four mathematics and therefore it has a low level coverage in the test. Geometry is a marginal topic in some projects, though in other projects it is considered to be quite important. The 10 percent geometry's representation in the test well reflects its present status in the mathematics curriculum for the primary school grades. Among the behavioral categories computation and comprehension received an equally high level of representation. Application, too, had a satisfactory level of representation.

It should be noted that the classification employed in Table 2.4 served the purpose of content validation only. The item classification scheme for the purpose of deriving subscores and

Table 2.4 Number of items in various behavior by content categories.

The behavior dimensions	Content areas				Total number of items
	Number System	Integers	Fractions	Geometry	
Computation		12	2	1	15
Comprehension	6	6	2	1	15
Application		5		2	7
Analysis		5			3
Total number of items	6	26	4	4	40

of presenting results was quite different. Details about that classification scheme appear in the next section.

Deriving Subscores

Though the major analyses reported in this study are related to the total score obtained by the students in the whole test, we also dealt with several subscores based on a selected set of the test items. For pragmatic reasons, we employed four classification schemes which were entirely orthogonal to each other and derived scores for items clustered together according to each of the classification schemes. Table 2.5 present the categories according to the four schemes. A brief explanation of these categories and the rationale behind them clarifies the reasons for dealing with them.

Topics - The list appearing in Table 2.5 here constitutes a slight modification of items classified in the grid of Table 2.4. This classification scheme was closer to the teachers' conception of topics in mathematics study and therefore preference was given to its utilization in the report.

Core-materials - The division of the items into a small subset of core-items and non-core items satisfied information needs related to the attainment of a minimum competency level in fourth grade mathematics by pupils in the nation's schools. The importance of such a distinction and the problems related to deciding what is "core" and what is beyond the core is discussed in the introductory section of this study, and will be more thoroughly treated in the chapter on results. It seems necessary to indicate here, however, that the definition of core items was done on the basis of analytical considerations by experts prior to the inspection of the test results. Nevertheless, after the test results had been summarized it was decided to re-define the core-items set. A detailed explanation of the reasons for redefining the core item set appears in the next section.

Year-of-teaching - The mathematics test used to examine the achievement level of students covered skills and topics taught in grade four, and also those which were taught at an earlier grade level. Mathematics is a hierarchically structured discipline and its teaching requires continuous and repeated drill and further development of concepts and understanding taught previously. Most mathematics books contain sections and exercises related to mate-

rials taught in previous years. There was, therefore, ample reason for deriving scores separately for topics and skills taught in grade three and four.

Table 2.5 Four schemes of item classification.

Topics	No. of items
Decimal System	3
Computation (Integers)	13
Word Problems	7
Fractions	4
Geometry	4

Core items	No. of items
Core	10
Not core	30

Year of teaching	No. of items
Grade 3	21
Grade 4	19

The Judgmental-Empirical Approach to Modifying the Minimum Competency Requirements

In the previous section we described the method of selecting a series of ten test-items to represent the Minimum Competency Requirements in fourth grade mathematics. As indicated, the ten items were identified on a judgmental basis. They covered various contents and skills prescribed by the syllabus and were judged to be easy items.

Nevertheless, after inspecting the results of the Minimum Competency items separately and the scores on the whole series together, the evaluation team expressed dissatisfaction. They felt that it would be appropriate to redefine the Minimum Competency Requirements by omitting three items which appeared slightly difficult, and to substitute them with easier items.

Table 2.6 contains a brief description of the topics represented by the items contained in the Minimum Competency series and the percent of correct responses in a sample of 2843 pupils, and it also contains information on the redefinition of the Minimum Competency Requirements.

It can be seen that the three items eliminated from the original set were not the most difficult ones. Considering only diffi-

Table 2.6 Items in the two Minimum Requirement Items Series

Specific item content	Item no.	Percent of correct responses	
		new series	old series
Number system	1	91.8	----------
Sequence of numbers	2	98.0	----------
Simple addition (2 and 3 digits)	3	94.8	94.8
Addition of three digits numbers	4	96.6	----------
Subtraction	5	89.9	89.9
Multiplication with numbers having 0s	6	79.4	79.4
Word problems	10	90.7	90.7
Single division	11	87.3	87.3
Simple multiplication	13	91.7	91.7
Simple word problem	16	89.0	89.0
Measurement	18	----------	81.1
Word problem of two steps	21	----------	75.9
Geometry	30	----------	76.0
	Mean	90.9	85.6
	SD	13.5	17.2
	Number of respondents	2843	

culty as a basis for eliminating items, one should eliminate item no. 6, representing a simple but computation-wise more complex item (2 x 4000 = _____; answered correctly by 79 percent of respondents), and not item no. 18 (which represents a measurement topic and had an easiness level of 81 percent).

Nevertheless, the team was led by content considerations, too. A comparative analysis of various programs revealed that geometry and measurement problems are not covered in all mathematics programs, and therefore it was decided to eliminate items from these content areas. Also the two-stage word problems seemed to be non-adequate for the Minimum Competency series.

One may criticize this approach to redefining the set of Minimum Competency series, and claim that this procedure leads to an unjustified augmentation of the observed scores on this set. On the other hand, it can be said that there are no commonly accepted rules with regard to item selection and setting standards for Minimum Competency Tests. Berk (1986) identified 38 different methods for setting standards on criterion referenced tests. In an attempt to classify these methods he defined a judgmental-empirical approach, of standard setting, which consists of combining judgmental and empirical considerations through giving greater weight to the judgmental considerations. Usually the operational procedure consists of two steps. The first one determines standards on a judgmental basis, and the second step adjusts these standards after inspecting the empirical data. Adopting the method labelled by Berk as judgmental-empirical, we changed three of the ten originally selected items. The significance of this modification is illustrated by the information contained in Table 2.7.

The figures appearing in Table 2.7 indicate that applying a severe standard (as in accepting scores of 9 and 10 to be indicators of mastery on the Minimum Competency Test), there were 77 percent who received passing scores on the modified test, versus 64 percent on the original one. Employing a more lenient standard of 7 (i.e. scores of 7 and 8 are also considered as passing scores) the corresponding figures are 95 percent and 89 percent. In other words, using the lenient standards on the modified test, 5 percent of the respondents failed to obtain the Minimum Requirements while on the original test 11 percent failed to do so.

The results obtained on the modified test seemed to be more compatible with results obtained in other investigations (see, for example, Razel et al, 1984). Thus the decision to use the modi-

Table 2.7 The frequency distribution of scores on the series of minimum tests

Score	Percent of pupils	
	new series	old series
up to 4	1	2
5	1	3
6	3	6
7	5	10
8	13	15
9	23	25
10	54	39
A11	100	100

fied version of the Minimum Competency Test was based on congruence between judgments of two types. First, that certain items should be eliminated from the test, and secondly, that a 5 percent failure constitutes a correct estimate of the failure parameters. The congruence of the empirical evidence with judgmental considerations provides support for the adequacy of the method employed here.

Nevertheless, for those who have to take responsibility for selecting items and setting standards a warning should be conveyed: any slight change either in the wording of the questions or in the setting of standards has serious consequences with regard to estimates made on the rate of failure.

Data Collection

According to the internal regulations of the Ministry of Education and Culture any large scale data collection in schools must be cleared by the Chief Scientist of the Ministry. Schools within

the country can be contacted only through the district offices of the Ministry of Education, and the head of the district office grant permission for data collection, only if the project was approved by the Chief Scientist.

Accordingly, the study plan and the instruments were submitted for approval to the Chief Scientist, and after obtaining it contact was made with the district offices to obtain local permission and cooperation. Schools that were drawn to the sample were notified by the district head of the Ministry and were instructed to cooperate with the evaluation study.

The questionnaires and tests were packed at the Israel Curriculum Center. A separate envelope was prepared for each school. The envelope contained the following items:

1. Instructions to carry out the examination.
2. Teacher's questionnaire.
3. Student's questionnaires and mathematics tests.
4. Student's answer sheets (responses to the two instruments were marked on a single sheet of paper; this arrangement saved the need to collate various sources of information for a single person).
5. Report form, which was to be filled out by the teacher and to be signed by the principal, indicating that the testing was satisfactorily carried out. On this sheet the teacher was asked to indicate the number of persons who were absent on the day of the examination. He/she was also asked to mark whether some irregular event occured during the administration of the test.

The addressed, sealed envelopes were deposited at the district offices of the Ministry of Education and Culture, and representatives of the schools were requested to pick up the envelopes from the office a few days before the day designated for testing. They were also asked to return the data and the test forms in the same envelope to the district office. Schools were allowed to keep a single copy of the test. The evaluation team obtained the data from the district offices.

The school principal was asked to open the envelope in the class at the time the testing started. First, the questionnaires were given to the students and they were asked to respond to the questions. It should be remembered that the same sheet of paper contained also the answer items for the test. After completing the questionnaire a short break followed. Pupils were asked to leave the answer sheet on their desks, and the teacher checked

whether all students marked their name on the answer sheets.

After returning to the class from the break, the pupils received the test forms and recorded their responses to the test items on the response sheet.

The teachers collected the materials, added to the answer sheet of each student a few biographical notes (the father's educational level, the pupil's latest school mark in mathematics, whether the pupil received a passing grade in the third grade minimum competency mathematics test organized by the Ministry of Education), and put all answer sheets into the envelope. The teacher also filled out the teacher's questionnaire and the report sheet about testing. Adding these two documents to the envelope, the schools took care to return the envelopes to the district office.

Out of the 100 schools drawn to the sample 99 returned the data. One school dropped out of the sample for technical reasons. Two schools were late in returning the data to the district offices, and therefore in some data summaries data from only 97 schools were used. Nevertheless, most summaries are based on the data collected in 99 schools.

In addition to the large scale data collection enterprise in 99 schools, supplementary data were also collected in a few judgmentally selected schools. This small scale data collection was personally supervised or carried out by the writer of this report. The following small scale data collection was carried out:

(1) Report on utilization - Six teachers were interviewed in their schools by the writer of the study to find out details about their utilization of the evaluation results.
(2) The writer of this report participated in several discussions led by members of the evaluation team with various stakeholder groups of mathematics teaching in grade four. Summaries of these discussions were recorded by the writer.

Data Analysis

The data were recorded on electronic floppy diskettes and processed by CDC (CYBER 855) Computer.

Two data files were prepared: a students' data file containing records for each student participating in the study; and a class level data file. The class level data file contained both aggregate data about the class (like percentage of students in the class characterized by a certain trait, such as passing the minimum competence subtest in mathematics, class means and standard deviations) and class related variables, such as school traits, teacher's traits and conditions of teaching mathematics (like the compensatory education budget index of the school, the teacher's training for teaching mathematics, and the textbook used in the school).

The data analysis was carried out through the SPSS data package (Nie et al, 1975) and through a Test Scoring Program (Lewy and Crawford 1968). Data summaries were made at individual level and aggregate "class and school" level. Since only one class from each school did participate in the sample there is an identity between the class and the school aggregate data level. Accordingly, the expressions class level and school level are used in the study interchangeably.

A variety of statistical procedures was used in the study, but the major summaries reported in the study are based on univariate and bivariate statistics.

Chapter Three
Mathematics Intended Curricula

As indicated in the section on the goals of the study (see page 40), an attempt was made to portray the intended mathematics curriculum for elementary schools in Israel. We will present, first, information about the National Syllabus and, secondly, about three widely-used instructional programs[1].

The National Syllabus

Teaching in Israeli schools is monitored by a centrally produced and universally adopted syllabus. It constitutes the official vehicle through which the nationwide curricula of all subject matters are communicated to interested parties.

The National Syllabus contains information about the educational goals and the particular curricular objectives schools do strive for, specifies the learning outcomes pupils are expected to achieve, suggests teaching-learning strategies that should be employed in classroom (or at least be considered by teachers in their day-to-day pedagogical work), and it also lists the subject contents pupils are supposed to study. The National Syllabus also contains didactical and methodological suggestions related to each particular content unit.

The National Syllabus of mathematics, which is currently used in Israeli schools, was introduced in 1972. It is an innovative curriculum. Developed in the second half of the 1960's, it was greatly affected by the "new mathematics" movements of the 1950's and the newly developed mathematics curricula of the 1960's in the

[1]The terms "instructional program" and "set of instructional materials" are used in this report interchangeably.

United States.

Indeed, the 1972 mathematics syllabus for the Israeli elementary school level emphasizes topics of application and problem solving rather than rote learning. It also prescribes the fostering of child's ability to create mathematical models suitable to concrete situations he/she actually encounters in daily life, and the enhancement of the child's ability to suggest reasonable solutions to problem situations presented. The new curriculum stressed learning the structure of the subject mathematics, which is supposed to facilitate the development of mathematics reasoning and comprehension. This curriculum subscribed to certain principles of Piaget's and Brunner's learning theories, as specified in the 1972 National Syllabus:

(a) The teaching of mathemathics moves from emphasizing ordinary and intuitive reasoning to logical and formal abstract thinking.
(b) Mathematics teaching for young children should use more basic and simple contents and should apply them to suitable rules and concepts. The child is potentially capable of shaping mathematic concepts, relations and rules even without knowing their conventional definitions. Therefore the teaching of any mathematics content should deal in some form with rules, relations, and concepts which will be later formalized by the pupil him/herself. By this mode of teaching the school equips the child to discover and apply laws, principles and effective ways of problem solving.
(c) Significant learning of mathematics implies that the child manifests knowledge and comprehension of the subject's structure;
(d) The most suitable instructional approach is based on manipulative learning, which means using concrete teaching aids which are either taken from the child's environment or specially constructed for the purpose of teaching.

The National Syllabus of 1972 broadened the scope of contents taught in elementary school by including topics not contained in the previous syllabus such as geometry, various types of measurement units, and principles of probability.

An additional curricular feature specified in the National Syllabus relates to a suggested time schedule for teaching the mathematics topics distributed across the six grades of elementary school.

With regard to the differential treatment of low and high-ability pupils, the 1972's National Syllabus does not provide any clear guidelines. It neither specifies which contents could be ommited in teaching low-ability pupils nor does it suggest alternate didactic and methodological activities for teaching low and high-achieving learners.

It should also be noted that the sequence of teaching various content units is not prescribed by the syllabus, and the teacher is informed that he/she enjoys the autonomy to determine both the sequence of the topics and which didactical modes he/she prefers.

The teaching topics for grade four, as listed in the national syllabus, are presented in Exhibit 3.1. Exhibit 3.1 also includes the suggested teaching time for the various mathemathics contents.

Widely-Used Instructional Programs

As already indicated above, the data collected in a sample of 272 schools revealed that 22 different mathematics instructional programs are in use in grade four of the elementary school in Israel.

Table 3.1 presents the distribution of most used mathematics instructional programs in the country.

Table 3.1 Instructional Programs Used in Classes

The program	Number of Classes	Percent of classes
One Two and... Three	173	31
Simple Math	125	22
New Mathematics	95	17
Etz-Hadar	56	10
The Discovery Method	29	5
Aurbach	24	4
And Another One	19	3
Other	41	8
Total	562	100

Teaching topics	Learning contents	Time schedule (in hours)
The Concept of Number	a. The decimal system of numbers based on the method of positions b. The Hebrew alphabet system (its numerical connotation). c. Review, The powers of 10 Better knowledge of the decimal system of powers. Knowledge of powers in shortened multiplication of equal factors. Reading and writing numerals above one million. d. The axis of natural numbers. e. Familiarity with negative numbers. f. Better knowledge of mathematical rules: the commutative law, the associative law, and the distributive law. g. The characteristics of 0 and 1 in the four basic arithmetic operations. h. The numbers and their factors (factorization, prime and factorizable numbers).	30
Sets	a. Sign of sets b. Equal sets c. Weighted sets	5
Mathematics operations with positive numbers	a. The four basic operations b. Problems (description of problems, 2-3 phases problems, using parenthesis, finding arithmetic means). c. Estimation and rounding. d. Problems related to time conception (the day and hour, calculating time periods during the day -in minutes-, the Hebrew calendar, the leap year in the Hebrew calendar). e. Diagrams Reading and drawing diagrams (horizontal and vertical lines). Block diagrams. Coordinate system.	45
Simple fractions	a. Fraction as operator, as descriptor of quantity, and as result of division. b. Different names to the same fraction. c. Expanding and simplifying. d. Addition and subtraction of fractions with equal denominator. e. Multiplying an integer by mixed number. f. Dividing a fraction by an integer. g. Problem solving requiring calculations with fractions.	25
Geometry	a. Parallel and rectangular lines. b. Drawing squares and parallelograms. c. The quadrilaters. d. Rotation. e. Symmetry. f. Transformations. g. Angles and their measurement. h. Measurement (of surface of quadrilaters, transformation of measurement units, estimation of length and surface). i. The box and the cube.	35
All		140

Exhibit 3.1 Grade four teaching contents listed in the 1972's national syllabus.

It can be seen that three programs have a wide distribution in the country: the "One Two and... Three" program, the "Simple Math" program, and the "New Mathematics" program.

In the following sections a more detailed description of these three programs will be given.

The "New Mathematics" Instructional Program

The "New Mathematics" instructional program was developed in the first half of the 1970s by Mr. Moshe Hershko, an experienced senior mathematics teacher in the Israeli elementary school system. He participated actively in the mathematics curriculum Committee which was established by the Pedagogical Secretariat of the Israel Ministry of Education and Culture for the production of a new mathematics curriculum. Mr. Hershko participated in the elaboration of the preliminary versions of the presently-used mathematics curriculum, in the formative studies carried out in schools to assess the adequacy and effectiveness of the new curriculum, and in the design of the final version of the 1972's national syllabus.

Mr. Hershko developed the "New Mathematics" instructional program based on the same epistemological and psychological principles and assumptions underlying the 1972 elementary school new mathematics curriculum. This program emphasizes the teaching-learning of mathematic structures, relations, and rules, and the basic concepts of the subject. The main instructional concern is to teach young children to know, comprehend, and use mathematical rules and concepts without employing the corresponding formal definitions. A basic postulate of the program is that the mathematical method must be applied for studying arithmetical operations. In other words, the object of instruction is the learning of mathematics structures.

The teaching-learning approach adopted by the program is shaped by Piaget's theory: the stage of intellectual functioning for pupils at the elementary school level is the stage of concrete operations. Therefore, the "manipulative" teaching-learning method is considered to be the most suitable. This instructional method is also considered adequate for the structural learning of mathematics. According to the "manipulative" teaching-learning method, pupils assume an active role in the instructional process, and they manipulate concrete objects which may or may not be related to

their natural environment. The manipulation of simplified models of objects or symbols, or small wooden bars (bdidim) are examples of using the "manipulative" instructional method. Concrete mate- rials of this type are the means for the fostering of meaningful learning. The assumption is that by manipulating suitable in- structional aids, pupils acquire the skills of learning by disco- very. In other words, pupils actively participating in the in- structional process, and discover mathematical rules and concepts. Oral and written drills will be introduced only after the pupils themselves have already manipulated concrete materials.

The textbooks cover the syllabus of all six grades of the ele- mentary school, and for each grade level a separate book was pro- duced. Teachers who use these books must design and create them- selves instructional aids and guide their use by pupils. The author also prepared Teacher's Guides which contain advice on how to create suitable instructional aids. Thus, both teachers and pupils actively participate in the teaching-learning process.

The author advises teachers to construct achievement tests and use them systematically to monitor the pupils' progress, and to identify strengths and weaknesses. Once weaknesses have been iden- tified corrective instructional activities should be provided immediately. Suggestions for test items for assessing pupils' achievement levels of the various content units are also provided in the teacher' guides.

The program was gradually implemented in the elementary school system. In 1975 it was used for first grade only, and later it moved forward year after year till it covered all grade levels of the elementary school. Teacher autonomy is relatively emphasized in this program. The teacher has to decide how to construct ade- quate instuctional aids, and he makes decisions about pacing instruction and determining the time for teaching each learning unit. He/she selects the most suitable test items for assessing pupils' achievements and also constructs the measuring instru- ments.

The Teacher's Guides for each grade level contain two lists of content units: contents listed in the 1972's national syllabus, and contents as they are sequentially dealt with in the textbook. A strict correspondence of teaching contents is observed between both lists. Nevertheless, the sequences differ from each other. Suggestions for time allocated to topics is also presented in the guide.

With regard to differential treatment of high and low achieving pupils, the program's author indicates in the teacher's guide what textbook pages may be skipped when teaching low achieving pupils. Teachers therefore have the autonomy to make this decision according to his/her own judgment, and in view of the pupils' levels of achievement. The textbook pages referred to as dispensable for teaching low achieving pupils contain the following teaching contents:

1. Sets, listing the elements, using signs like " "...
2. Moving on the axis of numbers in equal sized sets.
3. Multiplying and dividing by 10 (short way).
4. Factors and divisors.
5. Simple and complex word problems.
6. General form of subtraction.
7. Comparison of sums and remainders.
8. Rectangular lines.
9. Square of numbers.
10. Multiplication of three digit numbers.
11. Distances.
12. Weights.
13. Ways of carrying out division.
14. Problem of the type "how much is it more (or larger than).."
15. Calculating a part of a number, and the number on the basis of its part.
16. Rhombus.
17. Height of a triangle.
18. Scale of maps.
19. The volume of cube.
20. Negative numbers.
21. Operations with regular fractions.

The "One Two and... Three" Instructional Program

The "One Two and... Three" mathematics instructional program was developed in the second half of the 1970's by the Center of Educational Technology with active participation of the Israel Curriculum Center. This program embraces the curricula of all six grades in elementary school mathematics, and it was implemented gradually year after year since the early 1979. Its implementation was monitored through a small scale controlled experiment carried out in eleven classes of the Jerusalem and Haifa school

districts (Zemer and Raz, 1984).

The "One Two and... Three" program fits the 1972 innovative mathematics syllabus adopted in Israel for elementary school. This specific instructional program was also based on epistemological and psychological principles and assumptions related to those which guided the development of the national innovative mathematics curriculum.

Hence, the "One Two and... Three" program emphasizes comprehension and development of creative thinking instead of rote learning. The basic concern of the instructional process is with the use of mathematics rules, relations and concepts for teaching young pupils the most trivial mathematics topics. In other words, the structure of the subject mathematics constitutes the major objective of the teaching-learning process. The underlying psychological postulates are that children in the elementary level of schooling are still at the stage of concrete reasoning, and therefore they are in need of concrete representation of abstract mathematical concepts. In this sense, the "One Two and... Three" and the "New Mathemathics" programs are similar to each other.

The program uses a highly structured instructional method which is based mainly on concrete illustrative aids designed specially to represent abstract mathematical models. The illustrative aids are intended to constitute a comprehensive mathematical environ- ment (as distinct from the child's natural environment). These are deliberately constructed illustrative aids not taken from the environment the child intuitively knows, nor do they maintain any associative relationship with things beyond the mathematics context. In summary, the "One Two and... Three" program includes teaching of mathematical concepts by the use of "comprehensive illustrative aids", which are designed to serve as a proper context of mathemathics studies and as mathematical model. They are supposed to constitute the means by which the teaching-learning of all relevant mathematics topics of the elementary school curriculum will effectivelly be conducted.

The child's natural environment is dealt with only later, once the child has already learned how to use the concepts he/she learned intuitively and is able to effectively apply them for solving problem situations in his/her natural environment. For example, when a child has to deal with a mathematical problem directly related to his/her natural environment and then solve it, he/she should be capable of identifing the adequate arithmetic

procedures, deciding what suitable mathematics structures should be applied, and then correctly solving the problem.

This conception of dealing with the child's natural environment within the context of mathematics teaching-learning represents the basic difference between the "One Two and... Three" program and the "New mathematics" program.

The elements of the program consist of textbooks, Teacher's Guide, and a package of teaching aids containin additional worksheets, games, short stories and summative achievement tests. All the instructional materials are ready-made, and they were drawn from three sources: (1) instructional materials adapted from the Minnemast Program with permission of The Mathematical Science Foundation of the Minnesota University (USA); (2) instructional materials developed by the mathematics team of the Israel Curriculum Center; (3) instructional materials developed by the mathematics team of the Center of Educational Technology.

The textbooks constitute the principal teaching-learning device of the program. Eighteen textbooks were produced for the whole program, and they are designed for teaching mathematics in each of the six grades of the elementary school, i.e. the contents of each grade level arepresented in three textbooks. Each textbook consists of worksheets which include both presentation/explanation of new mathematical concepts, and a series of diversified drills which involve activities such as handwriting, drawing, painting, cutting and glueing. These worksheets are of four types: (1) explanation of new concepts children must learn; (2) drills of contents; (3) extension exercises marked with a bell which contain more complex assignments for high achievers; (4) review exercises of contents which were previously taught at a lower grade level and which are not related to the teaching topics being learned at the moment.

The games complement the drill activities, and they are designed for small group learning. Six games are included in the overall instructional program and they can be applied for a diversity of learning drills and at distinct teaching-learning levels. The games' contents vary according to the mathematics teaching topics, but the games themselves have fixed rules. It is the teacher's role to adapt the level of each game with the children's needs.

The short stories are not intended to teach new mathematical

concepts, but rather constitute a framework for instructional activities carried out after the new mathematical concepts were learned. Indeed, they are designed to incite interest, to motivate, and to serve as a model of how to learn mathematical concepts. The situations presented in these short stories are illustrated by pictures furnished as a supplementare teaching aid.

The achievement tests are essentially summative, and they are of two types: (1) partial summative tests used to assess the children's achievement levels ater the teaching of particular curriculum contents; and (2) a final summative test which must be applied at the end of each scholar year. They contain items intended to measure all curriculum-related contents taught at each given grade level.

The instructional approach is based on expository teaching for all pupils in a classroom. Hence, the teacher must play an active role in the process of teaching-learning. The teacher's autonomy about what instructional materials to use in class, and about what test items he/she whould use forassessing pupils performance is quite restricted. The teacher is supplied with all instructional materials (ready-made) the authors considered sufficient and necessary for promoting mathematics teaching- learning; i.e. the teacher does not have to construct or even organize any instructional aid for conducting his/her own teaching activities. Also, the teacher is not advised to construct and/or complementary or supplementary teaching materials. The achievement tests the teacher does use to assess learning outcomes are also ready-made and they constitute an integral part of the whole program. Teachers are not advised (at least explicitly) to construct their own achievement tests, and the possilibity that teachers may feel a need to conduct additional assessments of learners' actual performance is not even taken into account within the intended framework of the program.

Another feature of the program structure is the presentation of the list of mathematics teaching topics it includes. The whole set of mathematics at each given grade level are listed in the textbooks and Teacher's Guides. These mathematics topics are specified in terms of titles. The particular learning contents which describe operationally the topics are not listed however in the program's documents.

The teaching topics related to the matters of both arithmetic and geometry are dealt with within all three textbooks of grade

four (textbooks no. 10, 11, and 12), while the topic of measurement is included in one of them only (textbook number 12). Comparing the program's listing of teaching topics and the National Syllabus' listing of fourth grade teaching topics, one finds that they almost fully parallel.

The sequence of the topics listed in the program documents is not a rigid one. The teachers are allowed to determine the sequence of teaching contents they consider suitable to their pupils.

Theoretically, all pupils should learn all topics listed in the program. The documents do not contain any reference to topics which may be omitted in classes of low-achieving pupils. It does seem, though, that the authors anticipated that not all pupils will learn all topics. Indeed, teachers are clearly cautioned that pupils who did not learn in grade three either the topic "prime and decomposable numbers" or the topic "getting to know fractions" must learn them necessarily within the framework of grade four.

Simple Math: The Television Fourth-Grade Mathematics Program2

The Israel Instructional Television Center produced a mathematics program which has been widely used in Israeli schools. Unlike the previously described program, the Television Center developed a set of instructional materials for grade four-six only. The program consists of textbooks and a television broadcast series. A teacher's guide is in the process of preparation.

The "Simple Math" textbook series was constructed by subdividing the year-long curriculum into seventeen two-week learning units, each consisting of eight class hours. The main ideas of each unit in the textbook series provide the mathematical core of the corresponding telecast.

The videoprograms prepared by the Israel Instructional Television Center are: (1) Multiples and Divisibility; (2) Why Shouldn't You Divide by Zero; (3) Prime and Composite Numbers; (4) Geometry 1; (5) Factorization; (6) Powers; (7) The Decimal

2The description of this instructional program is based on the article "Simple Math" by Nisa H. Movshovitz (1983).

System; (8) Geometry 2; (9) The Arithmetical Operations; (10) Divisibility Tests; (11) Estimation and Rounding; (12) Geometry 3; (13) Long Division; (14) Negative Numbers; (15) Diagrams and Graphs; and (16) Geometry 4. Note that the grade four "Simple Math" program does not include the topic "Fractions" which is listed in the National Syllabus.

The main mathematical issues of the Telecast are: (1) definition of prime, positive and composite numbers, and the characteristics of the number 1; (2) the utilization of Erathothenes' sieve to "separate" prime from natural numbers.

The "New Math" program was also developed on the basis of the epistemological and psychological principles subscribed to by the 1972 National Syllabus. Indeed, with reference to the goals in the cognitive domain, the development of mathematical contents attempts to take into account the needs and limitations of the cognitive development level of the target population without relaxing the requirement of mathematical rigor. The unfolding of the program content proceeds from concrete experience and the examination of particular cases, based upon analogies where appropriate, through intuitive generalization resulting from an analysis of patterns and rules. As far as possible, new knowledge is not presented without an accompanying explanation. The intro- duction of new subject matter is carefully based upon reasons anchored in logical arguments. However, formal proofs are not presented. Mathematical language is kept as accurate as possible, only compromising when it interferes with the natural communi- cation ability of the target population.

The objectives of the program are defined as follows: (1) To motivate students to learn mathematics through dramatized children's stories; (2) To lower the level of fear of mathemat- ics; (3) To provide for the development of mathematical-logical reasoning, i.e. solid reasoning for basic facts by associating previous knowledge with new facts; (4) To facilitate retention of basic rules through the aid of musical rhymes; (5) To improve the assimilation of mathematical language into the natural speech of students; and (6) To introduce students to applications of mathe- matics in real-life situations.

To create a friendly atmosphere within the framework of the television broadcast series, thus aiming at the goals in the affective domain, the authors employed the following strategies: (1) There is no one mathematical authority who teaches. A young

actor whose hobby is math and excells in telling strange stories is the main adult character. (2) The cast includes children, allowing for identification of the audience with the screen action. (3) Girls do not play the traditional role of mathematical inferiors, and mathematics is not conveyed as a male role occupation. (4) Various segments of Israeli society are represented through the characteres, and any stereotypical role of mathematical failure is eliminated. "I catch on too" is the desired student reaction to be gathered from viewing the programs. (5) Humor is not neglected as a way of enhancing familiarity. The same purpose is served also by mime-acting, magic tricks, science fiction, dreams, sports, outings, parties and games.

The unique feature of the program is its relation to a story framework. The story framework is that of a young man, Danny, and two children who are friends of his. Danny, an actor in his professional life, is also very fond of mathematics. He spends time with the children at his home and takes them to different places where they share wonderful experiences. Whatever happens to them sparks Danny's mathematical associative imagination. On the spot he emerges with improvised mathematical stories. Two mime-a ctors, Tough and Luck, are the heroes of many of Danny's stories. Their "mathematical" antics contribute to making the study of mathematics enjoyable. The main concepts and rules of the telecast are expostulated by a special combination of film animation, vivid music and video technique. The central mathematical message of each issue in the telecast is put to rhymes and sung by Danny and the children. This is done in order to help the students memorize the main ideas.

Mathematical drama is a main feature of the Television program's video-series. Drama is assumed to be a good tool for serving both the affective and the cognitive mathematics objectives. The program presents mathematics as an enjoyable experience through drama which, in most cases, serves as an integral part of the treatment of the mathematical topic. The drama stems from surprising mathematical facts, patterns, or puzzles. Typical errors, known to occur commonly, are treated in the programs by forming a conflict- situation around them, or by bringing them to an absurd point. An example of mathematical triggers of drama is: the uniqueness of prime decomposition of a number is surprising, particularly in light of the numerous ways it can be decomposed when factors are not restricted to primes.

Teachers are advised to approach the program contents in class

according to the sequence appearing in the textbook. On the other hand, autonomy is granted to teachers in the construction of the achievement tests and their application.

Finally, it should be observed that the instructional approach used by the program is more approximate to the "New Mathematics" program than to the "One Two and... Three" program.

Chapter Four
The Implemented Curriculum

Results related to the implemented curriculum will be dealt with in terms of two major variable clusters: the conditions of implementation and the teacher's and learner's program related activities. In more detailed terms we will deal with the following variable groups: (1) the learners background; (2) the professional education of teachers; (3) the autonomy of the teacher; (4) the teacher's attitude toward the program he/she uses; (5) time related variables; (6) implementation at the classroom level; (7) opportunity to learn. Finally, we will introduce four scales which were used to measure teaching conditions in the classes.

The Background of Learners

Socio-economic variables - Two kinds of data were collected which contain some relevant information related to the socio-economic status of the pupils. One of them is a school-related variable and the other an individual variable. The school-related variable is the compensatory budget index. This variable was used as a stratifying variable for drawing the sample. Information about the compensatory budget index was available from the records of the Ministry of Education.

The second variable was the father's level of education. This information was supplied by the teacher, and was extracted from the school records. Thus no burden was imposed upon students to supply this piece of information. Fourth-grade pupils may not have precise information about their father's educational level, or they may be embarassed to divulge it. It seemed, therefore, more convenient and more valid to obtain this information from teachers.

The cross-tabulation of the two frequency distributions is presented in Table 4.1. The table contains only those pupils for whom information was available on both variables.

Table 4.1 Father's education according to the compensatory index of schools (percent)

Years of education	School's compensatory index			
	1 - 13	14 - 34	35 - 100	All
13 -	34	22	10	24
9 - 12	56	54	46	53
- 8	10	24	44	23
All	100.0	100.0	100.0	100.0
Number of pupils	934	756	504	2194

Data in Table 4.1 indicate that the fathers of 53 percent of the pupils in fourth-grade mathematics had studied 9-12 years in school, and they distributed quite similarly across the three levels of the school's compensatory budget index. On the other hand, the percentages of pupils whose fathers studied more than 12 years and of those pupils whose fathers studied less than 8 years are quite equivalent (24 percent and 23 percent). In comparing father's education level for pupils in schools with lower compensatory index (in the interval 1-13), it can be seen that in these schools, as expected, the percentage of pupils whose fathers studied more than 13 years is substantially higher (34) than the percentage of pupils whose fathers studied less than 8 years. The contrary can be observed in comparing father's education level for pupils in schools with higher compensatory index (in the interval 35-100): the percentage of pupils whose fathers studied more than 13 years is substantially lower (10) than the percentage of pupils

whose fathers studied more than 13 years. Yet, the percentages of these two groups of pupils are equivalent with each other in schools with a balanced percentage of pupils from high and low socio-economic strata (school's compensatory index in the interval 14-34).

These data reveal that, in Israel, pupils are differentially distributed in schools according to their socio-economic origin, and this fact may explain why the variance in the pupils' achievements are due rather to variance between-classes than to variance within-classes. In other words, it seems correct to say that in Israel, in general, the classrooms are quite homogeneous with regard to types of pupils.

Parents' interest in the child's studies - The parents' involvement in school life is considered a value in itself. It is the realization of the democratic idea that parents have right to determine what kind of education their children should receive. The parents' interest in their child's success in school is of concern both because it constitutes a link between life at school and the parents, and because it demonstrates for the child that the parents, too, attribute importance to his/her success.

The parents' interest in the child's school work and academic success has been extensively studied by researchers (Bloom, 1976). It was demonstrated that parents' active interest and their active helps in children's school work are powerful factors in determining the child's academic success. Wolf (1964) referred to the parent's interest in the child's school work as a process variable and gave evidence that this process variable explains a higher proportion of achievement variance than static home environment variables (such as parents' education). It should be noted that according to Davis (1981), the causal relationship between home environment and scholastic achievement has not been sufficiently substantiated by empirical evidence, nevertheless, due to the extensive literature about this topic and evidence about positive relationship between these variables it was decided to deal with this variable in the present study.

Therefore we posed questions to the pupils about their parents' interest in their work in mathematics, and in the significance their parents attribute to the child's success in mathematics.

Almost all (99 percent) of the students indicated that the parents attribute significance to success in mathematics study, but

only 24 percent of children said that their parents frequently speak with them about matters related to studying mathematics: 70 percent of the children said that their parents sometimes speak with them; and 6 percent said that their parents do not speak with them at all about studying mathematics.

The results indicated that studying mathematics is an important issue for the parents. Most of them speak with their children about this topic. It is difficult to say to what extent does the parents' speaking to their children provide support for, or, alternatively, raise fears concerning mathematics study, but since it is already a topic of interaction with children, schools should take initiative in guiding parents on how can they help their children to succeed in mathematics.

Utilization of pocket calculators - Calculators appeared on the market at the beginning of the 1970's, and since then have become a widely-used computational tool. Nevertheless, as indicated by Hutton (1985) no agreement has been reached among mathematics educators about the utility of calculators in elementary school mathematics, and textbook producers in the U.S. are cautious to incorporate the utilization of calculators in their programs. In the Federal Republic of Germany, the Ministry of Education has prohibited the use of calculators for computation in grades one to six (Hutton, 1985). In Israel, the Ministry of Education has not presented rules concerning their use in the school and has delegated the responsibility for making decision to the mathematics teachers. But in Israel, too, as in the U.S., producers of mathematics programs and textbooks refrained from taking a position concerning the utilization of calculators. The major mathematics programs examined in this study do not contain exercises requiring the utilization of calculators as an integral part of the program.

In light of such circumstances it was of interest to find out to what extent children use calculators at home and in school. Questions were directed both to pupils and to teachers concerning the utilization of pocket calculators in class. The responses given are contained in Table 4.2.

It should be noted that the correlation between using calculators and the socio-economic indices used in this study approaches zero, which suggests that using calculators is not a function of the financial status of the family.

The data summaries appearing in Table 4.2. indicate that in

Table 4.2. Utilization of calculators at home and in the school.

1. Question: Do you have a pocket calculator?

Responses	Percent of pupils
No, I do not have.	47
No, but in the school we do have	3
Yes, I do have.	50
All	100

2. Question: Do you use calculators?

Responses	Percent of pupils
No.	60
Yes, but only at home	21
Yes, but only in the class	5
Yes, both in class and at home	14
All	100

3. Question to the teacher:
Are pupils in your class allowed to use pocket calculators?

Responses	Percent of pupils
No.	45
Sometimes.	48
Always.	7
All	100

approximately half of the schools calculators were not allowed, and approximately half the children did not have experience using them. It seems that not only experts and teachers, but also parents could not agree upon the desirability of giving pocket calculators to pupils.

The Teacher

The Professional Education and Experience of the Teacher

Whether the intellectual capabilities of teachers and their professional education affects the quality of teaching has been examined in previous studies. Gage (1978) referred to them as presage variables. Getzels and Jackson (1963) summarized a large quantity of studies demonstrating that presage variables are highly indicative of success in teaching.

Both the teacher's professional education and his/her experience have been topics of study since the early days of scientific research in education.

In the present study, questions were directed to teachers about aspects of their professional education and their professional experience.

Basic professional education - Most primary school teachers in Israel receive their professional education in Teacher Training Institutes. A two-year professional education is considered to be the standard preparation for teaching. Teachers completing such programs are qualified teachers. Most Teacher Training Institutes also offer a third year of studies, and those who successfully complete this year have the qualification of Senior Teacher. Finally, some Institutes recently gained academic status and are entitled to offer a fourth year of studies for students who excell. Graduates of the fourth year gain an academically-recognized B.Ed. degree which, after some supplementary studies, allows entrance to studies for the Master of Arts degree at universities.

As indicated in Table 4.3, 30 percent of the fourth-grade teachers have a two-year Teacher Training Institute education, and 70 percent have a three-year professional education.

Table 4.3 Teacher's basic professional education level.

Teacher's basic professional education	Percentage of teachers			
	New Mathematics	One, Two and... Three	Simple Math	Total population*
Three years and university	67	68	68	70
Two years and other	33	32	32	30
Total Number of respondents	18	28	25	95

*It should be noted that "total population" in this Table and in the subsequent ones contains information about schools using all program variations in the system.

It is of interest to note that in a survey of science teaching in grade six (Zuzovski, 1987), it was found that 37 percent of the teachers had only a two-year Teacher Training Institute education. Additionally, more than 10 percent of all teachers had an academic degree obtained either from some university or from academically recognized Teacher Training Institutes (in Table 4.3 grouped with "three-year education").

Professional education in mathematics teaching - Teacher Training Institutes provide training for teaching mathematics in the elementary school for all teachers. The training program consists of a weekly two-hour period of lessons and workshops and has a three-year duration. The syllabus covers various topics taught in elementary schools including enrichment of subject matter knowledge and didactical principles related to various mathematics topics. The trainees receive some instruction in developing instructional materials and there is also a requirement for actual teaching under supervision of the Institute staff members.

Statistics are given in Table 4.4 about teachers participating in two types of programs: programs in Teacher Training Institutes, and special courses conducted by various mathematics projects operating in the country.

The data reveal that 57 percent of the mathematics teachers received some direct training for teaching mathematics. It is interesting to note that the "One, Two and... Three" program have

Table 4.4 Teacher's Professional Education in Mathematics.

Teacher's professional education in mathematics teaching	Percentage of teachers			
	New Mathematics	One, Two and... Three	Simple Math	Total population*
Participated in Teacher Training Institutes				
- based on textbooks actually used	6	35	29	57*
- on basis of different textbooks	44	19	48	
Participated in special programs in-service education of textbooks actually used	44	83	73	73
Total Number of respondents	18	28	25	95

*Total percentage of teachers who attended Teacher Training Institute

the highest percentage of teachers who received program specific education, while the "Simple Math" program has the highest percentage of teachers who obtained direct training for teaching mathematics.

The "New Mathematics" program, which is a private enterprise, has virtually no program specific training.

Program specific in-service education - Most curriculum projects have their own in-service education program. These programs are organized either directly by the projects and financed by them, or they are located at a Teacher Training Institute or university and financed by the In-Service Education Department of the Ministry of Education.

Educational programs of this type are provided either in the form of weekly meetings during the whole academic year (60 percent of such programs are presented in this format) or they are concentrated in full-day workshops on vacation days and on days of leave from routine teaching duties.

Eraut (1972) distinguishes between four in-service education

paradigms. A schematic representation of these paradigms appears in Tabe 4.5.

Table 4.5 A schematic representation of Eraut's (1972) four in-service education paradigms.

Goals	Stimulus	
	Internal	External
Improvement	1. Problem solving	2. Defect
Redirection	3. Growth	4. Change

The type of in-service training provided for users of a specific mathematics program represent paradigm number 4 of Eraut's scheme. They are initiated by an external agency, and frequently constitute a precondition for obtaining necessary instructional materials. Their goal is redirection rather than improvement of current teaching.

As indicated by the figures of Table 4.4 (see page 85) a very high proportion of teachers (73%) participated in such in-service activities. The highest participation in in-service programs is observed among the users of the "One, Two and... Three" program (83%).

Of the various topics included in in-service programs, teachers attach great importance to aspects of methods of teaching and to problems of teaching in heterogeneous classes. The enrichment of mathematical knowledge and comprehension of mathematical ideas and structures seems to less important to the participants. They feel that despite the importance of classroom management problems and issues of testing and diagnosis, these topics are not sufficiently taught at the in-service courses.

No systematic studies have been reported about the success of such in-service programs, though it may be assumed that innovative mathematics programs cannot be installed without teachers receiv-

ing specific program-related training. Thus, rather than being satisfied with the high proportion of teachers participating in in-service programs, one wonder how do those few get along who do not participate in these programs. Do they get sufficient support from on-going guidance, the teacher's guide, or some other source such as colleagues using the program? The data collected in this study do not provide answers to these questions.

The Autonomy of the Teacher

The innovative programs used in Israeli elementary schools are highly structured and they delegate little freedom to the teacher. Practically, the teacher is expected to follow a prescribed series of lessons.

While using the curriculum packages the teacher may device some supplementary curriculum materials, and frequently he/she is advised on how to proceed. The teacher makes a decision about pacing and determines the time devoted to curriculum units, though in the "Simple Math" program, if the teacher wants to incorporate the television tranmissions into the lesson even the pacing is dictated by external factors. Also, the teacher may decide to introduce review periods and rework units taught at previous grade levels. The major decision delegated to the teacher is that of selecting one of the programs available on the market. But in a great many cases the school does not allow teachers to make the selection, and he/she is not given autonomy.

The issue of teacher's autonomy should be an important determinant in successful mathematics teaching, as well as in other subject fields.

The importance of teacher's autonomy, or the teacher's participation in curricular decisions is strongly emphasized in contemporary literature. Cohen (1985) listed several reasons justifying the autonomy of teachers in curricular matters:

(a) Teachers best understand the abilities and interests of their own learners, so that curricular decisions are best made by teachers acting as professionals within the school.
(b) The teachers' involvement in decisions about curricula is an instance of democratic participation in decision making, one of the most cherished social values in our society.

(c) Teacher's involvement in decision making reduces external control on the teacher's work, increases job satisfaction, and creates commitment for implementing the program.

Silberstein (1985) expanded the meaning of autonomy in the context of school curricula and introduced the concept of the autonomous curriculum user. By this he also referred to the teacher who understands the unique features of a particular curriculum, identifies him/herself with its rationale, believes that it best fits the learners' needs, contributes his/her energy to the further development of the curriculum, and prepares supplementary materials to be used in the class.

The expanded definition of autonomy proposed by Silberstein (1985) is expressed in the Teacher's Questionnaire (TQ), where questions touching on program selection are proposed to the teacher. The distribution of responses to these questions are presented in Table 4.6.

The three questions related to teacher's autonomy deal with the actual practice (who makes the decision?), the ideal practice (who should make the decision), and the reasons for selecting a particular instructional program, a single textbook or other teaching material.

It should be noted that the school regulations issued by the Ministry of Education do not provide rules concerning the decision-making procedures related to the selection of instructional materials.

The data of Table 4.6 reveal that relatively few teachers (11 percent) decide on their own about the selection of an instructional material, but 22 percent of the teachers believe that it is desirable that the teacher make the selection.

The most common view is that the teacher, the principal and/or the subject coordinator should jointly make the decision. 55 percent of teachers hold this opinion, and in 53 percent of the schools this is the actual practice of instructional material selection.

Tying together all the responses in which either the teacher him/herself, or the teacher in unison with others, makes and should make decisions, we find a discrepancy between the actual

Table 4.6 Teacher's autonomy in deciding adopted instructional program.

Teacher's autonomy	Percentage of teachers			
	New Mathematics	One, Two and... Three	Simple Math	Total population
Decision-makers on adopted textbooks				
- the teacher decides	6	11	8	11
- the teacher, the principal and/or the coordinator together	67	41	56	53
- the principal and/or coordinator decide	16	41	20	28
- others (the teacher and others decide together or others decide)	11	7	16	8
Teachers' perception about who should decide about adopting textbook				
- the teacher decides	11	14	28	22
- the teacher, the principal and/or the coordinator together	44	61	52	55
- the principal and/or coodinator together	39	21	16	20
- others (the teacher and others decide together or others decide)	6	4	4	3
Teachers' reason for deciding in favor of given textbook				
- participation in in-service education	75	44	73	58
- dissatisfaction with previous textbook used	25	56	27	42
Total number of respondents	18	28	25	95

and the desired situation. There are cases (approximately 10 percent of the schools) in which teachers are not given the level of autonomy they believe should be given. Moreover, in 28 percent of the schools the decisions are made by persons who occupy a higher rank order than the teacher in the school hierarchy (i.e. the principal and/or the subject coordinator). Only 20 percent of the teachers believe that this practice is desirable, and 8 percent of the teachers feel that the principal and/or the subject coordinator unjustly disregard the right of the teacher to make decisions, or at least to participate in the decision making.

If we subscribe to the idea that teachers should participate in decisions about the instructional program, then the practice of decision making should be changed in approximately 30 percent of the schools in spite of the fact that approximately two thirds of the teachers in these schools are quite content to accept the authority and decision making of their superiors.

An interesting feature of the responses given to these questions is that the school supervisor is not considered to be a legitimate partner of the decision-making team dealing with school instructional programs. Decisions about instructional programs are considered to be an internal matter of the school. Even in cases where the teacher renounces his/her right to participate in the decision-making process, he/she still believes that the decision should be made at the school level, rather than by authorities operating outside the school.

The reasons specified by teachers for selecting a particular instructional program is another indicator of teacher autonomy in Israeli schools. Only 42 percent of the teachers indicate that they initiated the change due to dissatisfaction with the book previously used. The rest, 58 percent, confess that they were influenced by in-service training diffusion agents.

Attitude Toward the Program Used in the Class

The teacher's attitude toward the instructional program he/she uses in the class is an important determinant in the program's success. Formative evaluation studies put a great emphasis on the teacher's identification with the program rationale and his/her approval of various program components.

Bloom (1977) compiled a list of program dimensions which should

be evaluated by teachers in the formative stage of program production. His list contains items such as "Clarity and meaningfulness", "Level of difficulty", "Interest value for students", etc. He also suggest that teachers should provide an overall rating of the program examined. Chew Tow Yow (1977) distinguishes between two categories of criteria for judging the adequacy of instructional material: Instructional and Practical characteristics. In the first category are characteristics such as "Contents related to instructional objectives", "Interesting", "Self motivating", "Reinforcement to pupils, allowing immediate feedback", "The utilization of illustrations", etc. In the second group are items such as "Easiness to use (appropriate size of manupulative materials)", "Proper use of color proportion", "Durability" etc. Gall (1981) provides a more comprehensive set of criteria to judge the quality of instructional materials, grouping these criteria in four clusters: publication variables; physical properties; contents; instructional properties. In the category of physical properties she lists items like "Appropriate utilization of media" and in the category content she lists the question "Are the contents acceptable to teachers?". A strong emphasis on teacher's acceptance of the program appears in Silberstein's (1980) study. Silberstein contends that the teacher's acceptance of the program constitutes a precondition for its successful use. He presents a sequence of variables which constitute a "path" for determining program outcomes. At the first position of this path stands the teacher's familiarity with the program. The comprehension of the program is a critical issue mainly in programs which contain innovative elements of the subject taught. In innovative science and mathematics programs teachers frequently do not have a sufficient substantive knowledge of topics and ideas contained in the materials, and their participation in in-service training programs leave them still ignorant of important features of the program. Therefore the mastery of the program content by the teacher is a necessary condition for its appropriate use.

The second variable in this sequence or path, as suggested by Silberstein (1980) is the teacher's acceptance of the program or his/her identification with the program rationale. The identification has several dimensions, such as the selection of contents; the relative emphasis of various mental operations, as in memorizing facts, or learning computational rules; the teaching method; the sequence of activities, etc. This variable is extremely important in situations where conflicting views prevail among subject experts, and where alternative educational programs, representing different conceptions of teaching a subject, are

available. This is the case in mathematics instruction in general across the whole world. A great many competing programs run vigorous campaigns to promote the diffusion of their materials, and in Israel, too, the market is flooded with competing mathematics packages. Under such conditions it is important that teachers should understand the rationale of various programs and should adopt that one which best fits his/her conviction.

The third variable in Silberstein (1980) path is the actual implementation of the program. This variable will be discussed in the next section. In this section our focus is the teacher's acceptance of the program and his/her identification with it.

The summary of relevant information is provided in Table 4.7. The data touch on various aspects of satisfaction.

Appropriate difficulty level - The questionnaire provided three choice options for the teacher: difficult, easy, suitable. It was assumed that both difficult and easy books are not suitable for use in a class. After examining the teachers' responses to the question, it turned out that only 4 percent of the teachers selected the "easy" response, and therefore it was decided to collapse the two categories "easy" and "suitable" into a single category "suitable". This was done partly because of the low choice rate of the response "easy" and also because it became evident that the expression "easy" in most cases did not have a "negative" connotation. Examining the choice pattern it seems that the difficulty level of the mathematics books does not constitute problem for the users of the "One,Two and... Three" and the "Simple Math" programs. Across the whole sample 18 percent of the teachers complained that the materials are too difficult for the children. In contrast, about 40 percent of the teachers using the "New Mathematics" textbook complain that the book is too difficult.

Graphic elements - The books contain graphic elements such as graphs and illustrations, which have an explanatory role and are supposed to facilitate learning. It seems that the "New Mathematics" book have problems with its graphics. Only 33 percent of the respondents found these elements helpful, versus 79 percent in the "One,Two and... Three", and 70 percent in the "Simple Math" program.

Clarity of expanation - The clarity of the explanations is another area of complain. A high level of satisfaction is indicated

Table 4.7. Teacher's attitude towards the textbook used in teaching mathematics

Teacher;s attitude towards textbook in use	Percentage of Teachers			
	New Mathematics	One,Two and... Three	Simple Math	Total population
Difficulty level of learning method suggested in the textbook in relation to class pupils)				
- difficult	41	11	4	18
- suitable	59	89	96	82
Utility of textbook's graphs on the process of learning				
- neither help nor disturb	39	21	30	28
- help	33	79	70	63
- disturb and other	28	-	-	9
Clarity of textbook's instructions for pupils				
- sometimes they are not clear	82	57	50	65
- quite clear	18	43	50	35
Ranking of mathematics activities as the most and the less importants in the instructional process				
- memorizing facts				
- the most important	13	-	8	7
- the less important	63	60	25	43
- instruction of computation's techniques				
- the most important	31	36	8	21
- the less important	25	20	16	23
- word problem solving				
- the most important	6	8	16	14
- the less important	6	8	44	23
- comprehension of concepts				
- the most important	53	68	84	67
- the less important	6	4	4	7
Total number of respondents	18	28	25	95

by only 35 percent of the users in general, by 50 percent of the users of the "Simple Math" program, and by only 18 percent of the "New Mathematics" book.

Ranking outcomes of mathematics study - The innovative

mathematics programs emphasize the importance of comprehending mathematical concepts and structures and relatively less emphasis is put on recall of facts and on mastery of computational skills. Therefore the ranking of various types of outcome in mathematics study can be an indicator of program acceptance. The frequency distribution of importance attributed to four outcome types appears in Table 4.7 (see Page 93).

It can be seen that the users of all programs attributed a high significance to the comprehension of concepts and in this respect it can be said that teachers internalized well the rationale of the innovative programs. Computation skills occupy a low preference rank. The low ranking of word problems should not raise concern. It should be remembered that the Table 4.7 contains information about percentage of teachers who assigned the highest rank position to a particular outcome type. Though in the new mathematics program importance is attributed to word problems, by no means are they considered as the most important outcomes.

While one may observe some differences between the users of various programs in regard to the ranking of different outcome types, we are inclined to attach more importance to the similarity of ranking patterns across programs than to the slight differences observed between them.

In sum, the data should be seen as a general indication that teachers internalized the rationale of the new programs to a relatively high degree.

Time-Related Variables

The time devoted to studying a particular subject or topic is a decisive factor in determining the learning outcomes. Traditionally, the time devoted to teaching a particular subject is positively related to the importance attributed to that subject and to its perceived difficulty. In elementary schools across the world mathematics and language skills (or as labelled in English the three Rs -i.e. reading, writing and arithmetic) constitute the central core of teaching. These subjects are often referred to as "tool subjects", because they constitute a tool for academic studies in most areas of knowledge. In many educational systems there is no special time allocation for these subjects, and the teachers decide upon the time to be devoted to them on basis of

the progress made in the class to master prescribed curricular objectives. Priority is given to reading, writing and arithmetic, and the scope given to other subjects is dependent on attainments in these basic skills.

In the Israeli educational system the recommended time schedule prescribed for studying arithmetic in grade four is four weekly periods.

Contemporary educational research has extensively studied schooling time variables and its relationship to educational achievements (Berliner, 1979). In this context we will deal with allocated or scheduled time only. Problems related to the time devoted to various units within the program will be dealt with in a later section of this study in the context of dealing with the Opportunity to Learn (OTL) variable.

Through the Teacher's Questionnaire (TQ) information was collected about three aspects of time scheduling. The relevant information for the total population and for schools using the three most popular programs are presented in Table 4.8.

Periods in the weekly schedule - Most schools (62%) devote five periods per week to studying mathematics and among those who use the "Simple Math" program 77 percent allocate 5 hours per week. In the "New Mathematics" program group we also find one school where only three weekly periods are scheduled for mathematics. It is of interest to identify some characterists this school presents. The compensatory education budget index attributed to it equals 87, indicating that a high percentage of pupils are from low socio-economic strata, 14 pupils were in the grade four class participating in the present study, and they made use of a computer program package in mathematics teaching-learning of. This may explain why the teacher reported he devoted only three weekly periods to teaching mathematics in the classroom: it may well be that he/she did not take into account the weekly periods in which pupils made use of the computer. It should be noted that computers are also used in another school, where the teacher reported that three weekly periods were devoted to studying mathematics and a different curriculum program is used.

Time devoted to content review - Mathematics is a hierarchically structured subject and contents taught previously have to be constantly reviewed in the class. Mathematics programs prescribe reviewing exercises. These exercises have at least three func-

Table 4.8 Allocated time in fouth grade mathematics

Time variables	Percentage of Teachers			
	New Mathematics	One, Two and... Three	Simple Math	Total population
Periods in weekly schedule				
- three	6	-	-	3
- four	50	39	13	35
- five	44	61	77	62
Percent of class periods devoted to content review				
- 25 percent or less	71	78	88	78
- 25 to 50 percent	29	22	12	22
Time allocated for homework				
- 1 to 2 hours a week	56	42	30	33
- 3 hours a week	11	33	30	29
- more than 3 hours	33	25	40	38
Total number of respondents	18	28	25	95

tions. First they take care of the needs of children who have not mastered a certain topic. Secondly, they serve as a means of recalling what already has been mastered. Thirdly, in hierarchically structured subjects, reviewing old materials lead to a deeper understanding of what was learned previously. While review is contained in all arithmetic programs, and good teachers systematically allocate time to it, nevertheless, there are differences in the amount of time devoted to it in the schools. The importance of reviewing was proved by Butcher (1975) mainly for low and middle-ability children.

It should be noted that content review may be done both in the school within the framework of mathematics lesson and outside the

school as homework. Thus the information presented here does not fully reveal how much attention is given in a particular school to reviewing, but it provides a useful orientation to this important issue. The results reported here reveal that in schools using the three most popular programs, the percentage of teachers devoting more than 25% of class time to reviewing is greater than in other schools.

Homework - The Ministry of Education and Culture (1959) provided guidelines concerning the time to be allocated for homework. According to these guidelines fourth-grade pupils are supposed to have homework for not more than 45-60 minutes per day. In most schools fourth grade mathematics occupy not more than 1/5 of the total study periods. Still, pupils have more work in mathematics than in several other subjects and thus it is reasonable that they should do 1:30-2 hours mathematics homework per week.

While studies on homework failed to provide conclusive evidence about the impact of homework on various aspects of school life (Coulter, 1979), studies carried out in Israel have repeatedly shown that time devoted by an individual student to homework is negatively related to achievements (Padua, 1986). In the present study, too, the time allocated for homework has a small negative correlation($r = -.11$) with class means in mathematics achievement.

It should be remembered that the correlation describes the relationship between the teacher's estimation about time needed to make homework and class means. It may well be that at individual level the negative relationship between these two variables is even larger. "Weak" students need more time to prepare lessons than "strong" ones. This fact is frequently overlooked in setting policy for homework assignments.

Implementation of the Program at Classroom Level

Examining program implementation is considered to be one of the most important aspects of curriculum evaluation. Charters and Jones (1973) explain the observation that so many educational experiments end up with no significant differences between the experimental and control group by the fact that quite frequently there are no real differences between the treatments employed in the experimental and control groups. The experimental group may enthusiastically adopt an innovative idea and decide to implement

a certain treatment, but then it gradually returns to the routine way of operation and fully relinquishes the implementation of the innovation. In such cases, claim Charters and Jones, a "non-event" is evaluated, and therefore one should not be surprised to find out that no effects are observed. They wrote:

> *If the difference between what the research regards as "experimental" and "control" programs is fictional, the findings of the study will be fundamentally misleading irrespective of their nature. This logic, though simple, is often overlooked both in conducting and in interpreting evaluation studies.* (p. 5)

The same idea is emphasized by Scheier and Rezmovic (1983):

> *...the concept of 'degree of implementation' is critical in order to derive valid conclusions from both outcome and process studies of social and technological innovation. To correctly attribute the observed outcomes of a social program to the intervention, the researcher should have empirical evidence on the extent to which program components were implemented. Without such evidence, researchers may erroneously conclude that an intervention was ineffective when, in fact, treatment implementation was inadequate to afford a valid test of the program. Further, even succesful interventions may not be replicable without a knowledge of program components.* (p. 599)

Fullan and Pomfret (1977) listed four reasons why implementation should be studied. First, to better know what indeed has changed, and to avoid viewing the innovation as a "black box" without specific information about its actual parameters. Secondly, one must understand the reasons why so many educational changes fail to become established. It is important to find out which features of the innovation can be, and which cannot be realized in practice. Thirdly, without providing evidence that a program (or treatment) was adequately implemented it is useless to examine its effects, since failure to attain program goals may be due to the lack of adequate program implementation.

Curriculum experts tend to differentiate between determinants of implementation and the actual process of program implementation. Such distinction was first made by Fullan and Pomfret (1977). By "determinants" they refer to factors which may affect

the implementation, such as the complexity of the innovation, the availability of support needed to implement the program, etc. In contrast, by implementation they refer to the actual utilization of the program.

The clusters of variables discussed in previous sections, i.e. the time factors, the teachers' general and specific program related education, the teachers' autonomy for selecting programs, and their attitude towards programs used in the school represent the category, which has been termed by Fullan and Pomfret (1977) as determinants of implementation. In this section we will deal with variables which represent actual class activities, or are at least directly related to them.

In constructing an instrument for measuring the degree of implementation, we adopted Fullan's (1985) approach, that views implementation as a process of putting a change into practice, and perceives curricular changes as a process consisting of a set of separately measurable components. Therefore, measuring curriculum implementation implies identifying the salient components of the process, and examining to what degree each of these components is represented in the process of using the curriculum. According to this view, implementation is an umbrella concept. One may combine the measures of the separate components into a single variable, or alternatively use them separately. Measures of implementation are valid if they fulfil the following requirements: (1) the components specified represent salient features of the curricular change; (2) the data collected are directly related to the specified components; and (3) there is reasonable evidence about the veracity of the recorded data.

In our study we relied on questionnaire data and accepted the responses given by teachers as faithful description of their behavior. Accordingly, in providing evidence about the validity of the measures, we focused on the first two requirements specified above.

The implementation components examined in this study were derived partly from lists appearing in other implementation studies. Thus, for example, "durability", or as defined in the present study "the continuity of using a particular textbook" was taken from Fullan's (1981) study. Other components were identified through joint deliberations of the writer and team members of the Steering Committee. Thus, for example, the components "using the syllabus and the teacher's guide" was strongly emphasized by the

director of mathematics curriculum development team, and "organization of classes" was emphasized by the teachers participating in the Steering Committee. The final decision about selecting the implementation components was fully endorsed by the whole Steering Committee. Consensus among committee members representing clients of different types (experts in mathematics, curriculum development evaluation and teachers) was considered as a proof of the validity of the instrument. It should be added that the involvement of the Steering Committee members in selecting implementation components may fulfill an additional role, beyond establishing the validity of the measures, i.e. facilitating the utilization of the findings of the study. As Patton (1985) indicated, the participation of the clientele of a study in determining what kind of data should be collected enhances the utilization of the findings.

The second issue related to validity, i.e. the relatedness of the information to the selected components was also examined and approved by the Steering Committee. But since the components were defined in operational terms, the reader himself may serve as a judge of the validity. Due to the face validity of the questions used to elicit the data, the procedures described above seemed to be sufficient for establishing the validity of the measures. On the basis of analogy to what Hameyer (1985) calls procedural legitimation of curriculum (i.e. to establish the legitimation of a program by examining decision making rules and interaction processes), one could call the procedure described above as "procedural validation of a measurement instrument".

The seven components of implementation will be discussed in the present section according to their ascending proximity to actual classroom situations. We will start with those variables assumed to be weakly related to actual classroom situations and finish with discussing those variables more directly related to classroom situations. These variables are: "the continuity of using the actually adopted textbook"; "using the syllabus"; "using the Teacher's Guide"; "teachers' meetings dealing with mathematics instruction"; "receiving instructional guidance"; "using teaching aids"; and, "class organization for teaching".

Relevant data summaries are brought in Table 4.9.

The continuity of using the actually adopted instructional program - On-going use of a certain instructional program is by no means a garanttee it is well-used. In any event, experience with whole set of instructional materials across various grades may

Table 4.9. Teachers' actual implementation behaviors

Teacher's actual implementation related behaviors	Percentage of Teachers			
	New Mathematics	One, Two and... Three	Simple Math	Total population
Continuity of using actually adopting textbook:				
- did not teach other grades	22	32	36	31
- had used it in grades 1, 2, 3	17	18	-	12
- had used it in grades 2, 3	39	36	-	21
- had used it in grade 3	11	7	-	7
- did used different books in previous grades 1, 2 and/or 3	11	7	64	29
Knowing mathematics syllabus for grade four				
- do not know it	6	4	4	3
- know and do not use it	17	7	8	8
- kow and use it sometimes	22	14	28	31
- know and use it constantly	55	75	60	58
Using Teacher's Guide				
- do not use it	41	14	96	48
- use it for general planning of contents of teaching	41	50	-	33
- use it for planning part of the lessons	18	36	4	19
Participating on math teachers' meeting in the school				
- had not participated at all	17	39	46	37
- had participated once at year	28	19	21	19
- had participated more than one time	55	42	33	44
Receiving teaching guidance				
- don't receive teaching guidance	71	29	52	48
- receive from one of school's - teachers	17	14	20	17
- receive from a representant of adopted textbook who come to inside the school	6	25	8	13
- receive outside the school, together with teachers from different schools	6	32	20	22
Using teaching aids				
- working pages with exercises	100	89	88	88
- working cards	33	29	56	48
- games	22	43	32	35
- others	17	21	44	33
Teaching organization of classes				
- all pupils carry out similar activities	71	65	56	61
- homogeneous groups	23	12	8	12
- heterogeneous groups	-	19	16	11
- individualized work	6	-	8	10
- others	-	4	12	6
Total number of respondents	18	28	25	95

provide a rich experience and a better understanding of the program which will eventually lead to better implementation.

The data reveal that in the "New Mathematics" and "One,Two and ... Three" programs a large proportion of teachers used other books from the series which he/she uses in grade 4 (67 percent and 61 percent respectively), while in the "Simple Math" project none of the teachers had experience with previous volumes of the same textbook.

Using the syllabus - The syllabus contains a concise description of suitable teaching methods, and its repeated use may indicate the teacher's intent to implement the program as required by the syllabus, or to adapt it to the local conditions, while remaining faithful to its spirit. The data revealed that approximately 60 percent of the teachers use continuously the syllabus. Another 30 percent use it occasionally.

Using the Teacher's Guide - One of the innovative features of the new mathematics program is that they publish Teacher's Guide for each grade level. The Teacher's Guide contains direct instruction on how to use the instructional materials prepared for the pupils. They also contain information about the rationale of the program and its objectives, and provide recommendations concerning teaching strategies. Prior to the publication of the so-called innovative programs in Israel, in the late 1960s, the Teacher's Guides available on the Israeli market dealt with only the general features of teaching a particular subject. These guides were equally applicable to all textbooks used in the schools (see, for example, the book of Uri Blum). The innovative programs changed this situation and they produced Teacher's Guides which were related to the needs of those who used the materials of a particular program. "Using these guides" was considered in this study as indicator of program implemention.

The data contained in Table 4.9 indicate that there are great differences between various programs regarding the extent to which they use the Teacher's Guide. The "Simple Math" program did not publish a Teacher's Guide and so the users of this program do not have program-related guide to use. The most intensive utilization of a Teacher's Guide was observed among the users of the "One, Two and ... Three" program. But even in this group, only 36 percent of the teachers use the guides systematically for planning current lessons.

Teachers' conferences dealing with mathematics instruction - The items discussed above describe solitary teacher's behaviors, i.e. the teacher carries out these activities alone, without the involvement of other teachers from his/her school or from other schools. The participation in teachers' conferences is a public activity and it represents both a kind of public pressure and a public manifestation of identification with the school program. Moreover, the mere fact that the school has devoted a meeting to mathematics teaching is an indicator of a serious concern in the school for problems of teaching mathematics. It should be noted that, traditionally, most teachers' meetings or conferences have on their agenda administrative issues related to the current school life. In addition, topics emerge from the monthly circulars of the General Director of the Ministry of Education. The circulars prescribe actions, which require planning on behalf of the school and they need to be discussed by the teachers. The circular also prescribes special pedagogical tasks, such as introducing a study unit about current news of national interest, anniversaries related to outstanding events in the nation's life, etc. Naturally, the issues of grades, school discipline, and so on are also frequently on the agenda for teachers' meetings.

Dedicating a teachers' meeting to mathematics teaching, even if it is not a plenary conference and only those who teach mathematics participate, has not only instrumental value for improving mathematics teaching, but it also has a value-loaded significance. It puts matters of mathematics teaching onto the table of school concerns. It leads toward strengthening coordination of work at various grade levels and among various teachers, clarifies goals, strategies, etc., and therefore it has been included in the cluster of implementation variables.

The data in Table 4.9 reveal that in a very high proportion of schools (37 percent) the teachers did not participate in such meetings during the whole year, and one may assume that the school did not schedule them. In some mathematics programs ("One, Two and ... Three; and "Simple Math") it seems like as if guidance for teachers was an extra-school affair. Mathematics teachers are in direct contact with the program centers, and the scope of local decision making concerning mathematics teaching issues is quite limited. In contrast, users of the "New Mathematics" program have more school conferences dealing with mathematics teaching. This may be because the "New Mathematics" program does not have a currently-operating development center, and thus schools have to initiate activities related to implementing and improving mathe-

matics teaching, rather than rely on external support and guidance.

Receiving teaching guidance - In the Israeli educational system, up until the 1970's the major role of guiding teachers in their daily work was the responsibility of the school supervisor. In 1973, the Ministry of Education and Culture appointed a committee headed by the well-known educational psychologist, A.Minkovits to examine problems of school supervision in Israel and to suggest a role definition for supervisors. The committee recommended separating the evaluative roles from guidance roles. It recommended that the evaluation of teachers be done by supervisors, and that a body of experts be formed to fulfill the role of teacher guidance (Minkovits and Bashi, 1973). The rationale of these recommendations was that supervisors who evaluate the teacher's work, and whose evaluations may affect the teacher's professional promotion, cannot expect that teachers will frankly reveal to them the difficulties they encounter in their work, nor ask their guidance in coping with problems of teaching. As a result of these recommendations, a new position was created in the educational system, the position of teacher-guide (madrich). He/she does not fulfil supervisory roles and is not asked to write evaluation reports about teachers. After the establishment of the Israel Curriculum Center, the teams who developed innovative educational programs established their own guidance services, and started to provide special guidance for those who adopted their programs.

As a result of this, two guidance services operated within the system. One operates on behalf of the Pegagogical Secretariat, and deals with the whole teaching body of a group of schools; and the other operates on behalf of various innovative curriculum programs, and its work is limited to dealing with problems of teaching any one particular innovative program. Gradually the program-related guidance service became strenghtened and some of the innovative programs co-opted persons from the overall teachers' guidance service to their ranks. Nevertheless, these two types of teacher guidance services still operate within the Israeli educational system today. Eden (1979) provides a thorough description of the operation mode of the program-oriented guidance services in schools. He described how an innovative language arts program was implemented. To facilitate the implementation of this innovative program different patterns of mediation between curriculum writers and teachers were established. One of the patterns was that the subject coordinator of each school received special training from the team of the curriculum writers. The second pattern was that

subject matter consultants visited schools and guided the teachers in their work. A third pattern of mediation was a two phase guidance intervention. A small group of highly successful subject coordinators were released from their routine school duties for two days a week. One day per week they had to participate in in-service training programs, and the other day each of them provided further training for the school subject coordinators.

Similar patterns of consulting and guidance were established in mathematics, too.

Receiving guidance from experts who operate on behalf of the curriculum team involves a continuous effort on behalf of the teacher, parallel to his/her daily work in class. This enables the teacher to obtain help from the guidance service team on problems emerging in implementing a particular program. Therefore "receiving guidance" was defined as one aspect of program implementation.

The data presented in Table 4.9 (see Page 101) reveals that approximately half of the teachers in the whole system receive guidance in teaching mathematics. According to these data the most extensive guidance is given to the users of the "One, Two and ... Three" program, and in the second rank appears the "Simple Math" program. A surprisingly high proportion of users of the "New Mathematics" program do not receive guidance. The most common pattern of receiving guidance is the regional meeting of teachers, in which teachers from various schools in a certain geographic area receive guidance from an expert who operates on behalf of the curriculum program.

Preparing teaching aids - The preparation of teaching aids, such as working sheets, assignment cards, games, etc., is a sign of the involvement of the teacher in teaching the subject. By preparing teaching aids the teacher becomes a participant of curriculum development activities.

The results contained in Table 4.9 indicate that the teachers are active mainly in preparing working sheets with exercises. All users of the "New Mathematics" program are engaged in this activity, and 88 percent of teachers across the whole system carry out such work. 43 percent of teachers using the "One, Two and ... Three" materials and 32 percent of teachers using the "Simple Math" program also prepare educational games for use in the class within the framework of mathematics teaching. Although there is

no precise indication of the quality of the games prepared by the teachers, the effort they invest in such work constitutes a component of program implementation.

Organization of classes - Various mathematics programs used in grade four of the elementary school do not contain prescriptive specifications about the organizational setting of the class. One feature of the innovative programs is the minimizing of the so called expository approach of studying mathematics and they recommend that work within the class should be adapted to the individual needs of the learners. Despite two decades of using innovative programs in various study areas the Israeli school has not appreciable moved away from the expository approach to teaching, or as labelled in Israel, "frontal" teaching, where the teacher stands at the front of the class and 30-40 obedient pupils simultaneously carry out the work prescribed by the teacher. In the sample examined in this study, 60 percent of the teachers reported that they teach in this way. The high proportion of teachers in each curriculum program who employ the frontal teaching method is, indeed, a more significant than the slight differences observed between various programs.

Opportunity to Learn Mathematics Topics

The variable Opportunity to Learn (OTL) emerged recently as one of the important factors associated with educational achievement. Its research history is relatively short, though the rationale behind it is rooted in common sense pedagogy. As a matter of fact, the distinction between general ability or intelligence on one hand, and educational achievement on the other, is commonly made by considering the opportunity of a person to familiarize him/herself with the topic appearing in the test or to systematically study it. Intelligence tests contain questions which have not been systematically taught in school, while achievement tests contain questions which have been.

In the educational literature two different conceptualizations of OTL can be discerned. One of them was suggested by Carroll (1963), who defined ability in terms of time needed for a person to master a given skill. According to this conception, OTL should be expressed in terms of time devoted to learn a certain skill. Carroll distinguished between time engaged in studying something, and time needed to attain mastery of what was studied. In cor-

trast, Husen (1967) defined opportunity to learn in terms of the relationship between contents included in the curriculum and tests given to pupils. The concept defined by Carroll is an instructional variable, and that defined by Husen is a measurement variable. Wiley and Harnischfeger (1974) tried to create links between these two different definitions of OTL and introduced the term "time allocated to studying a curricular topic".

A great variety of methods have been used to measure the OTL variable. One of the simplest methods is keeping a record of weekly-study time periods allocated to a certain subject. Thus, for example, in an international comparative study the impact of weekly periods allocated to studying French as a foreign language upon achievement level in this subject was studied (Carroll,1977). Wiley and Harnischfeger (1974) measured the time devoted to studies by an individual on the basis of the school attendance records. The Beginning Teacher Evaluation Study (Fisher et al, 1980) used classroom observation techniques for measuring the time devoted to a particular topic.

In the IEA studies a questionnaire was administered to the teacher and he/she was asked to rate each item on an interval scale and to mark what percentage of the classroom pupils had an opportunity to study the topic represented by each particular item (Husen, 1967).

OTL measures were used in achievement surveys carried out in Israel for at least three purposes (Ortar, 1967; Lewy, 1972). First, to find out the actual coverage of prescribed educational objectives in schools. The study of Ortar (1967) first called attention to the discrepancy between the prescribed (or intended) curriculum and the implemented one. Secondly, there was an interest to find out the nature of the relationship between OTL measures and the cognitive achievements of the pupils. Thirdly, OTL measures were used in Israeli assessment studies as a means of reducing test anxiety both in teachers and in pupils (Lewy, 1983). In Israeli schools, the administration of centrally-produced assessment instruments usually create a feeling of anxiety. Both teachers and pupils are afraid that the tests may contain items which have not been adequately covered in the class. To alleviate this anxiety, the teachers and/or the students were asked to indicate whether they have had an opportunity to study in the school each particular item. In most cases it was indicated that the evaluation of the school and of each individual student will be done only on the basis of those items which were taught in the

school. A report about attempts to examine achievements in school based on information derived from OTL measures revealed that the relative position on a rank of a particular school in the total population is not affected by employing such a scoring technique. The rank of a particular school in the total population is the same when it is determined either by total test scores or by items which were marked as "taught in the school" (Lewy, 1972).

In the present study, two measures of OTL were employed. First, teachers were asked to mark each item which the pupils had an opportunity to learn in school. On the basis of the responses, an opportunity-to-learn score was calculated for each class. This OTL score was directly test-related. The second OTL score was curriculum-related. Based on the syllabus, a list of 21 educational objectives in mathematics for grades three-four was compiled. Teachers were asked to indicate whether the teaching in their class provided ample opportunity for the pupils to master each of these objectives.

The basic statistics for the two OTL are presented in Table 4.10.

The expectation that the two measures for opportunity-to-learn will converge was not met. Indeed, the relatively low correlation

Table 4.10 Statistics of the OTL scores.

Statistics	Test items related OTL	Curricular Topics related OTL
No. of items	40	21
Mean in percent	87.2	53.3
Standard Deviation	11.5	13.5
No. of responses	96	95
Range	21 - 40	6 - 18
Correlation	r = .26	

between these two measures suggests that the scales do not meet the requirements of convergent validity postulated by Campbell and Fisher (1959). Nevertheless, an analytical inspection of the two measures provides some explanation for the low correlation. Each of the 21 curricular topics represent a broad area of mathematical knowledge. Each of these topics in itself is a hierarchically structured domain, and therefore it may be represented by very easy items and also by more complex ones. It may well be that even in low achieving classes the teachers addressed themselves to all major areas, and taught some basic skills or items of knowledge pertaining to them. Thus, in these classes, a high OTL score was obtained on the curriculum-related OTL measure. At the same time they could obtain low scores on the test-items related OTL measure. In other words, the two measures examined different aspects of the variables OTL: the curriculum-oriented measure examined the breadth of the studies, and the test-items-related measure examined the depth.

Differences Between Schools

The frequency distribution of the responses to the two OTL measures are presented in Table 4.11. It should be kept in mind that in Table 4.11 the frequency distribution refers to the number of items or topics marked by schools.

It can be seen that approximately one third of the schools indicated that they taught almost all items included in the test. Approximately 80 percent of the schools taught more than 31 items included in the test. The "Simple Math" program is characterized by a lower compatibility to the test than other programs. Only 23% of schools studying according to that program taught most of the items, and in that group, in some instances, less than 26 items were covered in the class. A similar pattern of responses can be observed in the second part of Table 4.11 dealing with curricular topics.

Differences Between Topics

Information of another type derived from the OTL scores is the coverage of various test items and curricular topics in the school system.

Table 4.11 Frequency distribution of the two OTL scores.

a. test-items related OTL measure

No. of items studied	Number of Schools (in percent)			
	In the whole sample	In "New Mathematics" sample	In "One, Two, and ...Three" sample	In "Simple" "Math" sample
38 - 40	31.3	35.3	41.4	23.1
35 - 37	30.2	35.3	27.6	34.6
32 - 34	17.7	-	24.1	26.9
29 - 31	10.4	17.6	6.9	7.7
26 - 28	6.3	11.8	-	3.8
23 - 25	2.1	-	-	3.8
- 22	2.1	-	-	-
Total	96	17	29	26

b. curriculum-related OTL measure

No. of curricular topics studied	Number of Schools (in percent)			
	In the whole sample	In "New Mathematics" sample	In "One, Two, and ...Three" sample	In "Simple Math" sample
16 - 18	5.3	5.6	7.1	-
13 - 15	27.4	27.8	25.0	32.0
10 - 12	43.2	33.3	50.0	44.0
7 - 9	17.9	27.8	13.9	12.0
- 6	6.3	5.6	-	12.0
Total	95	18	28	25

Table 4.12 presents relevant information concerning the 21 curricular objectives listed in the teacher questionnaire. The percentage of schools providing an opportunity to master a particular objective is listed for the whole population and separately for those who studied according to the three different instructional programs most widely used.

Table 4.12 Opportunity to learn each of the 21 curricular topics.

The curricular Topic	Number of schools (in percent)			
	In the whole sample	In "New Mathematics" sample	In "One, Two, and ...Three" sample	In "Simple Math" sample
Getting to know big numbers	94.7	94.4	85.7	100.0
"Alef-Beit" system	81.1	72.2	71.4	100.0
Addition and Subtraction with numbers till 9999	94.7	100.0	96.4	88.0
Characteristics of "0" and "1"	90.5	94.4	82.1	100.0
Prime and decomposable numbers	80.0	66.7	96.4	96.0
Divisible signs by 3, 6, 9	72.6	66.7	85.7	76.0
Approximations	43.2	33.3	42.9	32.0
Progressive multiplication	89.5	94.4	92.9	84.0
Extended division	55.8	55.6	67.9	44.0
Word Problems	90.5	100.0	92.9	80.0
The number axis, addition of negative numbers and application; getting to know fractions.	26.3	33.3	21.4	8.0
Differents types of fractions, mixed fractions, addition and subtraction of fractions, fractions with equivalent or closing denominators, fractions multiplication.	25.3	11.1	64.3	-
Pocked calculators used for approximation exercises	13.7	5.6	10.7	20.0
Probability/Power	41.4	50.0	17.9	60.0
Review	42.2	50.0	32.1	36.0
Getting to know rectangles	8.4	-	3.6	24.0
The rectangles volume	4.2	-	7.1	8.0
Time (days, weeks, months)	62.1	50.0	67.9	56.0
Quadrilaterals family (including symmetry characteristics)	34.7	33.3	32.1	28.0
Geometry design	38.9	33.3	50.0	36.0
Symmetry	34.7	16.7	57.1	20.0
Total number of classes	95	18	28	25

One may divide the curricular topics into three groups: high coverage topics (more than 80 percent of schools provided opportunity to master the topic); medium coverage topics (30 - 79 percent of schools provided opportunity to master the topic); and, low coverage topics (less than 30 percent of schools provided opportunity to master the topic).

Relevant information is summarized in Table 4.13.

Table 4.13 Distribution of schools across levels of coverage of the curricular topics.

Level of coverage	Number of schools (in percent)			
	In the whole sample	In "New Mathematics" sample	In "One, Two, and... Three" sample	In "Simple Math" sample
High	3.2	-	3.6	-
Medium	90.5	94.4	96.4	88.0
Low	6.3	5.6	-	12.0
Number of cases	95	18	28	25

One may observe differences between schools grouped according to textbooks. It can be seen that the highest content coverage appears in the "One, Two, and ... Three" program, and the lowest one in the "Simple Math" program.

Among the topics which have low coverage in the whole system are "approximations" and "rectangular volume". It should be noted that the topic of "approximation" is highly emphasized in new mathematics programs, and their omission from the implemented program may require intervention by those who monitor and supervise mathematics teaching in schools.

Differences Between The Level Of Coverage Of Various Items

One may ask which items are heavely covered in the actual programs. This question approaches the issue of OTL with reference to test items, rather than with reference to schools.

Information about variations with regard to topics covered in schools can be derived from the summary of the OTL responses to the individual items.

Table 4.14 contains the frequency distribution of items according to the intensity of their coverage.

Table 4.14 Distribution of schools across levels of coverage of the test items.

Level of coverage	Number of schools (in percent)			
	In the whole sample	In "New Mathematics" sample	In "One, Two, and... Three" sample	In "Simple Math" sample
All schools	9	15	16	21
80 - 99	22	18	16	10
60 - 79	5	3	6	2
less than 60	4	4	2	7
Total	40	40	40	40

Comparision of the results obtained on the two OTL measures reveal that there is a higher congruence between the intended and implemented program using the items-related OTL measure than the one using the curriculum-related OTL measure. This finding can be explained by the fact that the test developers used informations concerning the actual coverage of mathematics topics in the system. Accordingly, they included in the test items which had reasonable coverage in the actual teaching practice. Despite the attempt to use items actually covered in schools, it turned that there is a considerable discrepancy between what is really covered in schools and what is contained in the curriculum. Considering the whole list of curricular topics one may discover even a higher level of discrepancy.

Items Covered in The Whole System.

It was of interest to find out which items of the test are most covered in the schools. As revealed in Table 4.14 nine items were taught in all schools. In Exhibit 4.1 we present two items which were covered in all schools, and for the sake of contrast in

Exhibit 4.2 we present two items which were covered in relatively few schools.

Comparing the items presented in Exhibits 4.1 and 4.2 it can be seen that while contents of word problem with integers was covered by all schools, high discrepancy exists between schools with regard to coverage of fractions.

Item no. 9	Each table has four legs. Dani counted 20 legs. How many tables did Dani see? Mark the suitable exercise for this problem. a) 20:4 b) 20-4 c) 40:10 d) 20+4
Item no. 16	How many books did 5 children buy, if each child bought 6 books? Mark the correct response for this problem. a) 36 b) 11 c) 22 d) 30

Exhibit 4.1. Contents of two items which were covered in all schools.

Item no. 26	Complete with the number that is missing in the square: $\frac{2}{3} = \frac{\square}{6}$ a) 2 b) 3 c) 5 d) 4
Item no. 35	A fence was built on a house roof. Its lenght is 50 meters, and its width is 30 meters. What is the perimeter of the whole roof? a)160 meters b)100 meters c)80 meters d)60 meters

Exhibit 4.2 Contents of two items which were covered in few schools (less than 60% of the schools)

Chapter Five
Attainment of Intended Instructional Goals

The results of the test provide information about the overall achievement level in mathematics of fourth-grade primary school pupils in the Israeli School System, and about achievements in specific subsections of the intended mathematics curriculum.

Those who are interested in the achievements of the educational system in general, such as the top management of the Ministry of Education, the parliamentary Educational Committee or the public in general, may focus their interest on the overall aspects of achievement in mathematics. For these "audiences" the mathematics achievements constitute one of the multiple desired outcomes of school learning, and they are likely to examine achievements in mathematics from two different points of view: (1) the relationship between intended and attained outcomes; and (2) achievements in mathematics in comparison with achievements in other subject areas.

The present study addresses itself to the first point of view specified above, i.e. the relationship between the intended and attained curricula. Issues related to the comparison of mathematics achievement to the achievements in other subject areas are outside the scope of this report. Nevertheless, it should be noted that the series of achievement surveys carried out by the Israel Curriculum Center provide comparative data which may be used for this purpose (Wolff, 1983; Lewy et al, 1983; Padua, 1986; Raz, 1986). Two sets of results will be of interest for the audiences mentioned above: findings related to the overall results on the test, and findings related to results on the Minimum Requirement sets. Assuming that items contained in the test represent the intended curriculum we may judge the success of the school system in mathematics on basis of the raw or percentage scores attained at the relevant test. This approach was employed by Minkovits et al (1977) with regard to total test

scores and also by most authorities who dealt with Minimim Competency Requirements in the U.S. (Gorth, 1979).

On the other hand, curriculum developers and teachers, who are dealing solely with mathematics, may be concerned both with the overall results and with diagnostic information based on scores attained in various subsections of the test.

Finally, results will be reported both at the individual and at the school/class level.

The summary of the individual scores contains information about the yield[1] of knowledge in mathematics in the whole target population. It constitutes an index of national importance, insofar as it predicts the proportion of persons possessing a certain level of knowledge in the future manpower of the country. It should be noted that in this section we deal with the summary of the individual scores, and our concern is national statistics of individual scores, rather than diagnostic information related to the individual per se. The utilization of the information obtained in this study for diagnostic purposes will be dealt with in the following section.

On the other hand, summaries of scores aggregated at school/class level provide informations which are relevant to the purposes of: school organization, grouping pupils in parallel classes, setting boundaries for the catchment area of a particular school, and so on.

Overall Results: Individual Scores

The information contained in the first two lines of Table 5.1 is pertinent to the overall achievements in mathematics in grade four of the Israeli school system. It can be seen that the mean score on the whole test containing 40 items is 74.5 percent of correct responses. This level of mean score appears to be satisfactory both from a technical and an educational point of view.

As to the technical point of view, it fits the requirements

[1]The expression "yield of knowledge" is used in the IEA studies to denote the overall quantity of knowledge acquired by learners (Walker, 1976).

made by experts in test theory that the recommended level of difficulty for multiple choice tests containing four possible answers items is 74 (Thorndike and Hagen, 1977; Lord, 1952).

Table 5.1 Means and standard deviations obtained on the total test and on subsections of it.

The Test	Number of items	On the Individual level			On the Class level			Reliability*
		Mean	Std.Dev.	N	Mean	Std.Dev.	N	
Total Test	40	74.5	17.2	2843	74.1	9.3	97	.88
Minimum Competency	10	90.9	13.6	2843	90.5	5.3	97	
3td grade	21	83.0	15.8	2843	82.6	7.1	97	
4th grade	19	65.0	21.4	2843	64.5	12.4	97	
Mathematical Computation	13	81.4	19.7	2843	80.8	10.2	97	
Mathematical Computation Grade 3	8	85.1	18.6	2843	84.6	7.8	97	
Mathematical Computation Grade 4	5	75.6	28.1	2843	74.9	16.1	97	
Word Problem	7	80.9	21.6	2843	80.4	8.4	97	
Word Problem Grade 3	4	87.6	21.3	2843	87.2	7.2	97	
Word Problem Grade 4	3	71.9	30.3	2843	71.3	11.5	97	
Decimal System	3	83.0	20.7	2843	82.6	7.7	97	
Geometry	3	53.4	30.3	2843	53.3	15.7	97	
Fractions	5	57.4	30.3	2843	57.1	19.6	97	
Measurement	2	70.5	34.1	2843	70.2	11.9	97	
Prime and Decomposable Numbers	2	63.2	41.6	2843	62.1	23.5	97	
Approximation, Divisibility sings, Basic Rules	5	73.7	24.4	2843	73.4	11.0	97	

* The reliability was calculated according to Kuder Richardson no.20.

As to the educational point of view, teachers in the Israeli educational system are used to consider the score of 70-80 as an

indicator of an acceptable level of achievement. This approach was adopted by Minkovits et al (1977) and it has a high level of appeal to Israeli educators.

In the measurement literature several attempts have been made to establish cutting-off points on an achievement scale, and to declare scores below that point as an indication of failure, and scores above it as an indicator of success. Glass (1978) provided a review of systematic attempts to establish cutting-off points and spelled out the limitations of various suggested methods. Moreover, he strongly denied the possibility of establishing a scientific basis for setting cutting-off scores and labelled all such attempts as pseudoscientific activities. He severely attacked attempts made by Mager to set standards as a part of defining an educational objective. He fully rejected educational objectives following the patterns suggested by Mager such as: "*The student must be able to spell correctly at least 80 percent of the words called out to him during an examination period*" (Mager,1962, p.44). Glass reacted to these attempts writing that

> ... *this language of performance standards is pseudo-quantification, a meaningless application of numbers to a question not prepared for quantitative analysis. A teacher, or psychologist, or linguist simply cannot set meaningful standards of performance for activities as imprecisely defined as "spelling correctly words called out during an examination period". Little headway is made toward a solution to the problem by specifying greater detail about how the questions, tasks or exercises will be constructed.* (Glass, 1978, p.238)

Despite Glass' criticism Haney and Madaus (1978) believed that one can make sense of what Glass called pseudo-quantification.

It is our approach that despite the reservations raised by measurement experts against the setting of arbitrary cut-off lines for determining passing scores, one cannot disregard an approach deeply rooted in the educational practice of a particular system. Moreover, since the tests were strictly curriculum-oriented, this practice also corresponds to "educational common sense".

The mean obtained on the Minimum Competency set of items, as expected, is higher than the mean of the whole test. Its actual value is 90.9 percent of correct answers indicating a high level of achievement.

Complementary information to that presented in Table 5.1 appears in Table 5.2.

Table 5.2 contains the frequency distribution of the scores. It can be seen that 15.9 percent of the students had a score higher than 90 percent of correct responses on the whole test, and on the Minimum Competency set 53.7 percent of the students had a perfect score (i.e. answered correctly all 10 questions).

It was found useful to establish different cutting-off lines

Table 5.2 Frequency Distributions of the Total Test Percentage Scores and of the Minimum Competency Set Percentage Scores.

Intervals	Total Test		Minimum Competency Set	
	At the individual level	At the class level	At the individual level	At the class level
91 - 100	15.9	3.1	53.7	63.9
81 - 90	25.6	26.8	23.4	34.0
71 - 80	22.8	36.1	12.6	1.0
61 - 70	15.1	29.9	4.7	1.0
51 - 60	10.1	3.1	2.7	-
41 - 50	5.7	1.0	1.5	-
31 - 40	2.5	-	.7	-
21 - 30	1.6	-	.3	-
11 - 20	.6	-	.2	-
0 - 10	.1	-	.2	-
70% and above it	-	-	94.4	96.0
above 66%	73.0	76.0	-	-
number of units	2843	97	2843	97

for determining the passing score on the whole test and on the Mininmum Competency set. For the whole test the cut-off line was determined as correct response to 2/3 of the test items. In the present study a 40-item test was used and, accordingly, 27 correct responses were defined as a passing score. For the Minimum Competency set, the passing score was set at a slightly higher level, i.e. more than 70 percent correct responses. In practice this score is a necessary upward rounding of the 2/3 correct responses.

The results presented in Table 5.2 indicate that 64.5 percent of the pupils passed the requirements set for the total test, and 89.7 percent of the pupils passed for the Minimum Competency set. Leaving an allowance for the score of 70, as well, the passing rate is 95%.

One may summarize these findings in the following way: 64.5 percent of the students obtained a satisfactory score on the test, and an additional 25.2 percent of the pupils failed the regular test for grade 4, but passed the Minimum Competency test, and 5.6 percent of the pupils failed even according to the lenient criterion of the Minimum Competency test.

Overall Results: School Mean Scores

Tables 5.1 and 5.2 (see Pages 117 and 119) contain informations about the distribution of school/class mean scores across 97 schools which participated in the study. As expected, the distribution of the school mean score has a more limited range and, accordingly, a smaller standard deviation than the individual scores have. This is indeed an artifact of dealing with aggregate scores rather than with individual scores. Even in the "weakest" school there are students whose scores are higher than the lowest individual score in "high achieving" classes, and, therefore, a school mean is necessarily higher than the weakest individual score. This principle also applies to the highest school mean score. It is necessarily below the highest individual score. The differences between the upper limit individual score and the upper limit school mean scores are not well observed due to the utilization of relatively large score intervals, but the highest individual score was 100.0 percent and the highest school mean score was 92.9 percent.

The proportion of the individual and school standard deviations

approaches the ratio of 2:1, i.e. the individual standard deviations are twice as large as the aggregate standard deviations.

It should be noted that the proportion of the score variation between schools to the score variation within schools is a highly significant characteristic of a school system.

A full equality between schools in an educational system would mean that there is no variation between the school means on any measure of achievement. All schools, or at least most of them, would have very similar mean achievement scores, and variation among scores would appear mainly within schools. Such an ideal situation of achievement distribution is quite unusual even in countries which strive for a high level of social equality. Nevertheless, schools in Sweden, for example, are characterized by a score distribution of this type (Walker, 1976). The variation between schools is relatively very low, and the source of score variation is wholly located within schools. In such school systems the standard deviation in a single school is almost as large as the standard deviation in the total population. On the other hand, in developing countries the variation between schools is extremely high and quite frequently the highest scores in low-achieving schools are below the lowest score in high-achieving schools.

The variation between schools in the Israeli school system is relatively high. It may be of interest to compare the distribution of achievement scores in grade four mathematics in two schools, one of which represents the high achievement and the other the low achievement schools in the system.

A graphic presentation of the two score distributions appears in Figure 5.1.

It can be seen that the overlap of scores is relatively very low. School 1 has a spread of scores starting from 39 up to 100, and school 2 has a more limited range of scores. The majority of children in school 2 has a more limited range of scores. The majority of children in school 1 have a lower score than the lowest scores in school 2.

A more precise expression of the between school variation is the correlation ratio, or η^2 (etha square). The etha squares across countries reported for achievement in reading in the IEA is presented in Table 5.3.

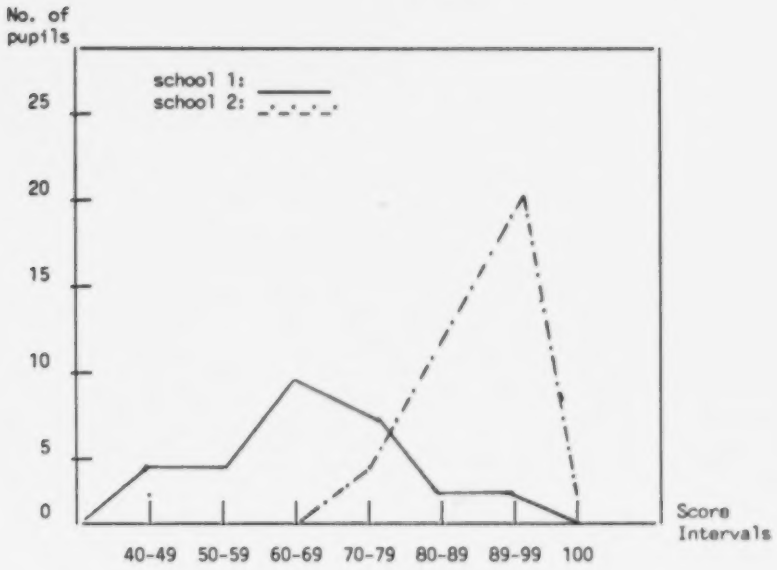


Figure 5.1 Distribution of fourth-grade mathematics scores in two schools.

Table 5.3 Homogeneity of Achievements Across Schools in Fourteen Countries

The Country	Pupils aged 10
Italy	53
Iran	29
England	20
United States	26
Flemish Belgium	31
French Belgium	26
India	82
Holand	28
Hungary	31
Israel	36
New Zealand	-
Scotland	21
Chile	62
Sweden	15

Source: Lewy et al (1978, p. 47)

As spelled out by Lewy et al (1978) the data in Table 5.3 indicate that in Israel, more than in a majority of developed countries, the schools have quite homogeneous populations, i.e. the variation between schools is relatively high.

In the present study, we found that $\eta^2 = .26$. This suggests that in the last decade equality between schools has slightly increased. This can be considered a positive development, since, at least in this respect, the Israeli statistics came closer to those observed in many industrialized countries. In practical terms it means that the differences between very good schools and very bad schools became lower than they were previously.

The Distribution of Low Achieving Students Across Classes

Since one of the major concerns of the educational system in Israel is the low achievement level of a substantial proportion of the school population and the distribution of these low achieving pupils across schools, we focussed our attention on those pupils who failed to pass the Minimum Competency Test. As already indicated in Table 5.2, 10.3 percent of the sample failed to pass the intermediate standard set for this test. It was of interest to examine the distribution of this 10.3 percent of the students across schools. There are two reasons that these distribution statistics were examined. First, one may assume that teachers are not able to deal with more than 2-3 low achieving pupils in one class. If the low achieving students would be equally distributed across classes there would be no great burden on a single teacher to deal with too many severely handicapped pupils. Secondly, the overall score distribution in classes along with the number of severely retarded pupils in the class constitutes one of the most significant environmental variables which affect school learning. A low achieving pupil who studies in a class accomodating high achieving pupils is more motivated to study, and usually, also obtains better instruction than a pupil of an equal achievement level who studies in a poor class environment (Coleman, 1966).

Table 5.4 contains relevant information. The data reveal that in 14 schools (approximately 15 percent of all schools) there were more than 3 pupils who failed the Minimum Competency test, and in two schools there were at least six pupils who failed. On

the other hand, in 30 schools there were no pupils who failed this test.

Table 5.4 Distribution of low achievers across classes

On the Minimum Competency Set		On the Total Test	
Number of pupils by class	Number of classes	Number of pupils by class	Number of classes
0	30	0	4
1	26	1 - 2	14
2	13	3 - 4	12
3	14	5 - 6	12
4	9	7 - 8	14
5	3	9 - 10	9
6	1	11 - 12	15
7	-	13 - 14	5
8	1	15 - 16	8
Total	97	17 - 18	-
		19 - 20	3
		21 - 22	1
		Total	97

The frequency distribution of Table 5.4 provides evidence that schools in Israel do not have equal proportions of low achieving pupils, and that the burden of dealing with the bulk of those pupils who fail the Minimum Competency test falls on a small proportion of the schools.

Results Related to the Subsections of the Test

The previous section of this chapter dealt with the Minimum Competency items included in the mathematics test. Beyond classifying the test items into core (or Minimum Competency) and non-core items, three additional classification principles were employed: the school grade level a particular item was taught; word problem versus numerical exercises; and the specific topic of the item.

The results related to item groups according to the three

classification schemes are presented in Table 5.1 (see Page 117).

The prescribed grade level of topics

The test used in this study examined the achievement level in all topics taught in the school till grade four. Topics taught in grades one-two seemed to be extremely easy for pupils in grade four, and consequently, the test items covered knowledge taught in grades three and four. In other words, it covered cumulative knowledge attained in grade one-four. Separate scores were produced for items appearing in the curriculum of grade three and for those appearing only in the curriculum of grade four. Since mathematics teaching follows a highly structured hierarchical sequence of teaching, and exercises related to topics taught in previous years are systematically interwoven into "new" study materials, it is expected that pupils will demonstrate a higher level of mastery on these items than on "new" items taught in grade four only.

Lewy and Chen (1977) have demonstrated that in subjects like geography, history, etc., pupils tend to forget topics they learned in the years before taking an achievement test. In contrast, in seems that in mathematics, reading comprehension, and comprehension of a foreign language, the knowledge is cumulative, and students have a greater mastery of knowledge for items appearing in the curriculum of lower grades than of those included in the curriculum of the year of the examination.

For this reason it was decided to derive separate scores for the grade three and grade four curriculum items included in the test. Relevant information appears in Table 5.1 (see Page 117).

It can be seen that the achievements on the grade three items are considerably higher (the mean is 83 percent of correct responses) than on grade four items (the mean is 65 percent of correct responses).

One could bring up two possible explanations for this discrepancy. First, there is the possibility that in some schools the teachers have not yet fully covered the content of the grade four curriculum and thus items appearing in the grade four section of the test have not yet been taught in these schools. The second possible explanation is that a large group of learners need a long time for acquiring mastery of a new topic in mathematics. Only

after repeated corrective instruction in the class, and repeated exercises, are the learners able to master a complex topic. Their knowledge of the topic immediately after their first encounter with it is meager. The data contained in Table 5.1 do not help us know whether these explanations possess an acceptable level of validity, or which one has a higher level of validity.

Word Problems versus Numerical Exercises

The distinction between solving verbal prolems and carrying out numerical calculations is well known in the field of mathematics. The importance of verbal problems is strongly emphasized in recent mathematics curricula. Indeed, the utilization of mathematics in daily life, as well as in the field of research and development entails solving verbal problems. Recent position papers by professional organizations have strongly supported "making applications" a major focus of mathematics teaching (National Council of Teachers of Mathematics, 1980). Lesh (1980) indicates that the ability to use elementary mathematics concepts in everyday situations is one of the most important skills required for mathematics, and its teaching should start at the lowest grades of primary school. Nesher (1980) pointed out that word problems included in mathematics textbooks should deal with realistic applications, and that important, everyday types of problems should be adequately represented in the textbooks. Quite frequently, word problems are considered to be more difficult than numerical calculation exercises.

The importance placed on word problems in the new curricula motivated us to calculate separate statistics for them and for numerical exercises.

To eliminate the danger of confounding the effects of "grade level the item was first introduced" (as discussed in the previous section) on the type of question, it was decided to present the results of word problem and numerical calculations separately for curriculum contents of grade three and of grade four.

Data in Table 5.1 (see Page 117) reveal that with respect to grade three-related items, the achievement on mathematical computation is lower (85.1 percent of correct responses "in mean") than the achievement on word problems (87.6 percent of correct responses "in mean"), while concerning grade four-related items the mean achievement on mathematical computation is higher than the

mean achievement on word problems (75.6 against 71.9 percent of correct responses). The difference in favor of achievement on word problems on grade three-related items may be because in grade four more intensive review is devoted to this topic than to techniques of mathematical computation. This is perhaps due to the general feeling that solving word problems is a complex and important content of mathematics subject matter. Anyway, both on grade three-related items and on grade-four related items there is no considerable difference between the means of achievement obtained in mathematical computation and word problems. This finding suggests ("all other things being equal") that in Israeli elementary school mathematics the "mean" difficulty related to learning of computational techniques is at least equivalent to the "mean" difficulty related to the learning of word-problem solving.

On the other hand, however, on grade three-related items the standard deviations of class averages on mathematical computation and word problems practically equal each other (7.8 and 7.2). On grade four-related items the variation between-classes of achievements on mathematical computation is quite higher (standard deviation of 16.1) than the variation between-classes of achievements on word problems (standard deviation of 11.5). A tentative explanation for this could be that teachers differ strongly in regard to the relative importance and the degree of difficulty they attribute to new topics of mathematical computation taught for the first time in an advanced grade level. Accordingly, large between-class differences may appear with respect to school periods dedicated to this topic and the amount of repeated corrective teaching in the classes. On the other hand, a low variation between classes appears to exist on teaching-learning emphases placed upon new types of word problems taught at the same grade four level.

This finding seems to indicate that one important source of achievement variations in the Israeli elementary school mathematics is variation between classes. In other words, one can assume that both the differential distribution of low achieving pupils across schools and the differential importance attributed to curriculum contents by teachers seem to explain the heterogeneity of the Israeli elementary school level.

Specific Topics

In addition to the major topics described above, scores were

derived for item clusters repesenting the following topics: decimal system (3 items); geometry (2 items); measurement (3 items); fractions (5 items); "prime and decomposable numbers" (2 items); "basic rules, divisibility signs and aproximation" (5 items).

The results obtained on these subsections of the test are reported in Table 5.1 (see Page 117). It can be seen that the mean achievement obtained on the decimal system set is considerably higher (83 correct responses) than the means achievement obtained on the other four sets of items, while the mean achievements on fractions and geometry are the lowest ones (57.4 and 53.4). It is reasonable to assume that the topic decimal system is more efficiently mastered than the others because a knowledge and comprehension of it form a basis for efficiently learning advanced computation and word problems. In fact, the mean achievement on the decimal system is quite similar to the mean achievement on the whole set of mathematical computations in grades three and four, and on the whole set of word problem items. It should be noted, also, that a relatively high correlation exists between class means on decimal system items and on mathematical computation items on one hand, and decimal system items and word problem items on the other (both coeffecients are .64). The low means on fractions and on geometry are probably due to the fact that at the time the test was administered a considerable proportion of classes had not yet studied these topics, and the majority of classes in which they were taught had not yet had sufficient time to practice them.

Results Related to the Individual Items

The smallest subunits of a test on which scores can be produced is the individual item. In most tests the items are clustered into subsections representing a well-defined educational objective or a specific content unit. The grouping of test items into clusters entails decisions which represent personal biases. Therefore, in several large scale achievement surveys the researchers avoided creating scores on subsections of the test and contented themselves with providing item statistics. This practice has been followed in the National Assessment of Educational Progress Surveys (Tyler, 1985), and also in the Van Leer Study of Achievements in the Arab Educational Subsystem of Israel (Bashi et al, 1981).

While in our survey we have not adopted the approach of focus-

ing on item statistics only, nevertheless we found that looking at results obtained on individual items may provide useful insights for the decision maker.

In the next section we will present illustrative examples about the utility of such analyses.

The Test Items Known By Almost All Pupils

The test contained 17 items which more than 80 percent of the pupils answered correctly. Among these were six items which more than 90 percent of the pupils answered correctly.

The easiest item was item no. 2 (Exhibit 5.1), which more than 98 percent of the respondents answered correctly. In 56 schools none of the respondents failed to answer this item correctly.

- Which number follows 599?
a) 590 b) 600 c) 598 d) 601

Exhibit 5.1 Content of item no. 2

Item no. 4 (Exhibit 5.2) was also very easy, which 96 percent of the students answered correctly, and in 39 schools none of the respondents failed the item. Moreover, in 86 schools less than 10 percent of the respondents failed to mark the correct response.

- Add the two numbers and mark the correct answer for this exercise		2037 + 506	
a) 543	b) 2543	c) 2533	d) 2536

Exhibit 5.2 Content of item no. 4

These items represent knowledge possessed by the total population.

The Test Items Known By Few Pupils Only

Some items of the test were answered correctly by relatively few learners.

Thus, for example, to item no. 25 (Exhibit 5.3) only 30 percent of the respondents gave the correct response, i.e. barely more than a correct response by chance.

- Add and mark the correct response for the exercise:

$$\frac{1}{2} + \frac{1}{4} =$$

a) $\frac{1}{6}$ b) $\frac{3}{4}$ c) $\frac{2}{6}$ d) $\frac{2}{4}$

Exhibit 5.3 Content of item no. 25

Discrimination Among Schools Versus Discrimination Among Individuals

One of the most instructive features of the item statistics produced in the present study is the possibility of identifying items discriminating between schools as well as items which have little discrimination power between schools and yet do discriminate between individuals. Lewy (1974) called attention to the importance of these two different kinds of discrimination, and pointed out that in curriculum evaluation the clientele is more interested in discrimination between schools than in discrimination between individuals. He suggested using the standard deviation of the class means as an index of discrimination among classes, i.e., a relatively high standard deviation should be considered as a sign of high discrimination between schools.

This approach to item analysis was adopted by several researchers (see Madaus and Araisian, 1978), and in this study it was found useful to present such pairs of items which are similar to each other, with regard to their overall level of difficulty, but which also reflect different educational realities. One of the items in each pair on Table 5.5 has a high level of discrimination power between schools and the other has a relatively low discrimination power between schools.

As the data of Table 5.5 reveal, on item 28 four schools had less than 20 percent correct responses and in 27 schools the percentage of correct responses was at the interval of 21-40.

It can be seen, additionally, by directly inspecting the mean of correct responses given to each item and the correspondent standard deviations of the classes' success levels, that item 29 did not differ substantially from item 28 with regard to their level of difficulty (58 percent and 59 percent of correct responses), but the frequency distribution of success levels across schools is highly different for the two items. School differences are relatively high for item 29. This can be seen by directly inspecting the frequency distributions contained in Table 5.5, and also by looking at the standard deviation of the class averages. This statistics is 22.1 for item 29 versus 17.1 for item 28. Similar observations may be made with regard to the other two pairs of items, i.e., in each of them the two items did not differ significantly from each other with regard to the corresponding levels of difficulty, but they did differ in the respective frequency distributions of the levels of difficulty across schools.

This kind of data enables one to point out whether differential effectiveness of classes on specific learning objectives are observable, and to find out on what topics classes differ from each other in the effectiveness of instruction. In looking at the content of the items presented in Table 5.5 it may be seen that a differential effectiveness of classes exists regarding teaching of the topic "prime and decomposable items" which was measured by items number 29 and 32, and also regarding the topic "fractions" as measured by item number 24 (the respective standard deviations are 22.1, 25.9, and 25.5). Also, it can be seen that "applying basic rules" (as represented by item 17) is more difficult (59.5 percent of correct responses) than "solving computational exercises involving basic rules" (72.6 percent of correct responses on item number 8), though the standard deviations of the two items

Table 5.5 Number of schools with different levels of success in pairs of items

Item no.	Item content	Percent of correct responses in the schools					Mean	Std.D.	
		- 20	21-40	41-60	61-80	81-100			
28	Which of the following numbers is four hundred and eight thousand three hundred and twenty two? a) 480,322 b) 408,322 c) 48,322 d) 4,008,322	4	27	37	28	1	58.8	17.1	>pair 1
29	Mark the number that divides into 3 without remainder. a) 155 b) 143 c) 171 d) 229	12	18	33	27	7	59.8	22.1	
8	What number is missing in the square? $5 \times 31 = 5 \times 30 + 5 \times \square$ a) 11 b) 0 c) 5 d) 1	-	11	27	44	15	72.6	16.6	
24	Subtract and mark the correct response for the exercise: $\frac{7}{9} - \frac{2}{9} =$ a) $\frac{5}{6}$ b) $\frac{9}{4}$ c) $\frac{2}{6}$ d) $\frac{2}{4}$	10	9	11	38	29	73.2	25.5	>pair 2
17	Rami bought 5 envelopes each one with 5 stamps, and Yosi bought 20 stamps. How many stamps did Rami and Yosi buy? Mark the exercise that answers this question. a) 3 x 5 + 20 b) 3 + 5 + 20 c) 5 + 20 d) (5 + 20) x 3	5	21	41	29	1	59.5	16.4	>pair 3
32	Which of these numbers is a prime number? a) 27 b) 31 c) 45 d) 33	16	9	26	39	7	61.7	25.9	

are quite similar (16.4 and 16.4). Further evidence that "applying basic rules" is a more problematic topic in grade four than "solving numerical exercises involving basic rules" is given by the frequency distributions of the items' difficulty levels: on the former item, five schools had less than 20 percent correct responses, in 41 schools the percentage of correct responses was the interval of 41-60, and only one school had more than 80 percent correct responses; on the other item a different pattern of distribution is observed: there was not one school having less than 20 percent correct responses, 44 schools had correct responses of the range of 61-80 percent, and in 15 schools the percent of correct responses was at the interval of 80-100.

How much mathematics do fourth grade pupils know

In the previous sections attempts were made to examine the mastery of core-requirements through scores based on a prespecified set of items. In other words, the items contained in the core (or minimum requirement) subtest were carefully selected and scrutinized by experts. The status of an individual and of schools and subgroups was described in terms of mean scores on the core-subtest and percentage of students passing an arbritrarily-set pass level on this subtest.

The summaries of these data reveal how well the pupils master contents and skills which the assessment experts considered to be the core-requirement.

But one may approach the problem by dealing with core-requirement in a less structured way, too. One may decide not to specify core-requirements in advance. Instead, one may examine results related to various subtest scores and to test items, and determine a-posteriori whether they can be considered as satisfactory. Some a-posteriori considerations appeared in the Van Leer study for setting core-requirements (Minkovits et al, 1977). The researchers in that study drew items for the core-subtest after observing the item analysis data, and selected only such items which were correctly answered by a relatively high proportion of respondents. In a similar manner, such a-posteriori considerations played an important role in establishing the final version of the core-requirement subtest in the present study. Nevertheless, both in the Van Leer study and in the core-requirement series of the present study, the summary was based

on the results related to a prespecified set of items.

In this study we ventured to make an additional step toward a non-preordinate' method of dealing with core-requirements.

In addition to answering the question whether students know what has been defined as a core-requirement, we also asked the open question what do students know and to what extent can this knowledge be considered as satisfactory. Such an open approach to dealing with educational achievement appears in Ganor's (1983) study, who examined achievement in history studies in a school for culturally disadvantaged students. She justified this approach by claiming that one cannot know in advance what kind of skills and contents have been taught in a particular class, and therefore it may be more appropriate to examine what students learned in a class, and then to judge the adequacy of what they know.

We judged that it was appropriate to examine the adequacy of the learners' knowledge both at the whole system level and at school level. Since the smallest unit exposed to curricular and instructional variation is the class, it was found that examination at the individual level is not appropriate.

The straightfoward approach of examining what pupils have mastered is to examine the distribution of the item difficulty indices.

Table 5.6 contains relevant information at the whole system level and at class level, and represents a broad variety of schools sampled.

The data presented in Table 5.6 indicate how much learners know. It provides quantitative information only. In order to be able to interpret this information, one needs to complement it with qualitative information.

Using a severe criteria of 80 percent as a criterion for satisfactory level of achievement (i.e. 80 percent of the students mastered a particular content or skill), one may observe that in the whole sample 42 percent of the test items represented such units of knowledge. Schools greatly differed from each other in this respect. For example, in school no. 74, 94 percent of the items

---

'We borrowed the expression non-preordinate from Stake's model of responsive evaluation (Stake, 1972).

Table 5.6 Percentage of test items difficulty at the whole system level and at class level.

Percent of students responding correctly (difficulty indices)	Total population	School's number							
		1	14	18	31	55	74	76	90
100	-	7	-	3	-	-	25	10	30
90 - 99	17	15	3	25	27	3	62	45	17
80 - 89	25	30	20	13	33	-	7	25	33
	42	52	23	41	60	3	94	80	80
70 - 79	25	20	17	10	10	15	3	11	7
60 - 69	20	5	15	17	17	5	-	3	3
50 - 59	3	10	15	15	3	27	-	3	3
- 49	10	13	30	17	10	50	3	3	7
All	100	100	100	100	100	100	100	100	100
Average percent of correct responses	74	74	60	69	77	46	92	87	85

were mastered by the students at this level, while in school no. 55 only 3 percent of the items were mastered at this same level.

Viewing the test results in this way, reveals interesting features of achievement, which have been previously overlooked in routine analyses of test scores.

The presentation mode appearing in Table 5.6 contains a kind of information which has not been treated in conventional summaries of achievement surveys. Assuming that the test fairly represents curricular units the pupils were supposed to study, Table 5.6 reveals what percentage of these units have been mastered by a particular group (class, school, etc.).

While most frequently one will find a high level or positive correlation between the average correct responses in a class and the number of items mastered by the members of that class (as it is the case in the data appearing in Table 5.6), nevertheless, such relationship is not a necessary one. One can imagine a situation that a class obtained an average of, say, 40 percent of cor-

rect responses in a way of having 40 percent correct responses to each one of the test items. In that case none of the curricular units represented by the items was mastered by the class. Alternatively, an average of 40 percent correct responses in a class can be otained by having answered 50 percent of the items by 80 percent of the students and by having produced wrong answers by all individuals to all other items. Such cases may be extremely rare but, at least mathematically, they may occur. Therefore the mere information about group average together with the commonly resported univariate statistics do not reveal anything about the proportion of curricular themes mastered by the group.

To illustrate the unique information obtained from Table 5.6 let us refer to the facts that in school no. 55, only 3 percent of the curricular units represented in the test have been mastered by the class (i.e. one of the 40 items), and in school no. 31, approximately 40 percent of the curricular units represented in the test have been mastered. Such information may be useful for monitoring the teaching, since it provides information for the teacher as to what requires further treatment in the class.

The mode of presentation exhibited in Table 5.6 may be especially useful in educational systems that have not centrally prescribed a core curriculum for all students and have given autonomy to the teachers deciding what to teach in their classes. Under such circumstances, it may be unfair to examine attainment in minimum competency on the basis of a centrally designed test. It may be more useful to examine what has been attained by the students than to examine whether a fixed subset of items has been mastered.

Achievement Profiles

The term achievement profile denotes the characterization of a person, or of a group of persons, on a set of related variables (Keeves, 1985). In educational practice the score vector of a person in various school subjects constitutes a measure of his/her achievement profile. Recently, in the British educational system the term **educational profile** has been used for "*...records which describe outcomes of education in the form of comprehensive statement referring to the range of a pupil's educational experiences, competencies, and interests*" (Kant and Orr, 1985). Such profiles contain information not only on the cognitive outcomes of

education but also on areas of social and personal skills, perseverance, punctuality, etc.

While in most cases the notion of achievement profile is applied to measures on a broad scope of characteristics, one may also develop achievement profiles within the framework of a narrowly-defined domain of desired outcomes of schooling. Thus, for example, within the area of a single school subject one may produce an achievement profile containing scores on various curricular objectives of that subject.

In the present study a series of scores were derived from the test, representing major subareas of the mathematics curriculum in grades three-four.

The following subareas of mathematics achievements were included in the study: word problems grade three, word problems grade four, mathematical computations grade three, mathematical computations grade four, fractions, the decimal system, and geometry. The full profile discussed in this chapter consists of 7 subarea scores.

Profiles based on separate scores for subareas of the curriculum provide useful information of a diagnostic nature. It reveals the relative strength and weakness of a person or of a group, and it may therefore serve as a basis for planning and implementing corrective action.

A hypothetical example of achievement profiles is presented in Exhibit 5.4. The scores appearing in this Exhibit may represent the characteristics of two persons, two schools or two groups of any size.

The subjects represented by the two profiles have an average score of 60 on the three subareas together, but they differ greatly from each other with regard to the configuration of their scores on the three topics represented in the Exhibit 5.4.

It can be easily seen that the profile provides information which is obliterated if only total test scores are reported.

Nunnally (1967) called attention to three major informational features of profiles: level, dispersion and space. Level refers to the height of the score on each element contained in the profile. In the example of Exhibit 5.4 one may observe that on topic

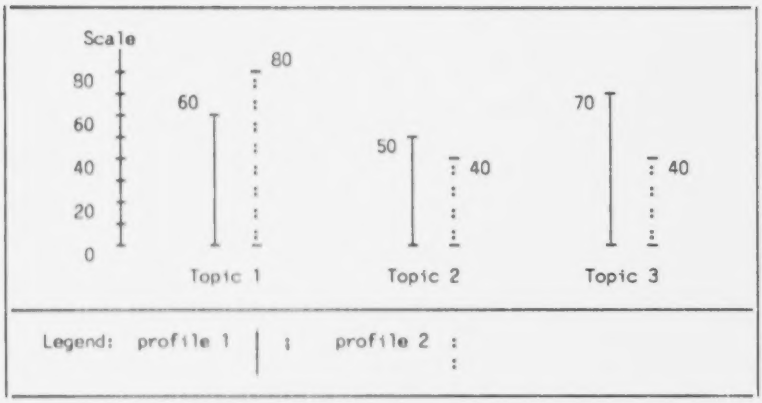


Exhibit 5.4. A hypothetical example of two achievement profiles.

1, and only on topic 1, are the scores in both profiles above 50. The dispersion refers to the distance among various scores contained in a particular profile. In the example above one may calculate the distances between scores 1 and 2; 1 and 3; 2 and 3. The sum of these distances for profile 1 is 40. For profile 2 the distance is 80. Therefore one may say that the dispersion of profile 2 is larger than that of profile 1[2].

Finally the third feature, called by Nunnally space, provides information about the rank order of scores in each profile. It can be seen that in the profile 1 the highest score appears for topic 3, while in profile 2 it appears for topic 2.

In the present section we will present summaries concerning the three above-mentioned features of the achievement profiles. We will deal with the achievement profile of the whole population, of sub-populations using a particular textbook, of individual schools and of single persons within a school.

Achievement Profile of the Whole Population

The purpose of dealing with achievement profiles in this study

[2]It should be noted that the simple distance between all score pairs is only one way of measuring the dispersion of the scores.

is to illustrate the possibilities of using such profiles as a starting point of discussing implications of the test results with teachers and administrators at various levels of the educational hierarchy. The information provided in this section, as it is, may have meager informational value about the state of art in teaching fourth grade mathematics without supplementing it with ideas, explanations, suggestions generated by decision-makers as a reaction to inspecting the achievement profiles of individuals or groups, whose education is of their personal responsibility. But since the purpose of this report is exploring the potential of using evaluation results for decision making, we present here profiles which may illustrate their potential use as input in various decision contexts.

Information about the achievements on the sub-scales of the test is contained in Table 5.1. The sub-scales appearing in Table 5.1 are not fully independent of each other, since various subscales contain common items. Thus, for example, the score Minimum Competencies contains items which appear in other scales, or the scale "Word Problems Grade 3" is fully contained in the "Word Problems" scale as well as in the Grade 3 scale.

In this section we will deal with subscales which do not have common items.

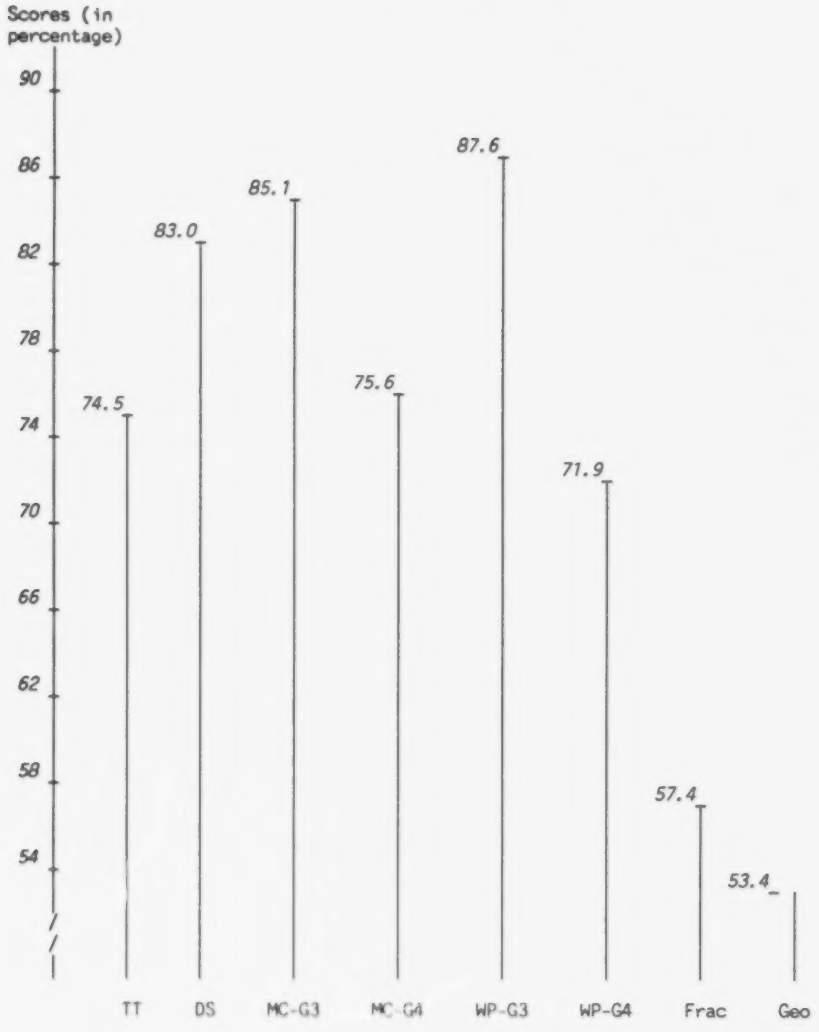
The achievement profile of the whole population is presented in Graph 5.1 in terms of the percentage scores.

The information contained in this profile may serve basis for decisions about curricular and instructional issues in the whole system.

It can be seen that the scores on Geometry and Fraction are relatively low, and that Mathematical Computation Grade 3 and Word Problems Grade 3 are higher than the corresponding scores for grade four. The figures appearing in this profile and their pattern may serve as a basis for interpreting subgroup or individual profiles.

Profiles of Users of Various Textbooks

The three major textbooks used in the system differ from each other with regard to their structure emphasis and sequence of topics. It was of interest to examine whether there are differ-



Legend: TT - Total test
DS - Decimal System
MC-G3 - Mathematical Computations Grade 3
MC-G4 - Mathematical Computations Grade 4
WP-G3 - Word Problems Grade 3
WP-G4 - Word Problems Grade 4
Frac - Fractions
Geo - Geometry

Graph 5.1 Achievement profiles on seven subscales of the whole population

ences in the achievement profiles of the users of these three textbooks series.

In examining the achievement profiles of subgroups of the population as well as of individual schools and individual pupils we will use T scores, which are standardized transformations of the raw scores[3].

By using T scores one eliminates the impact of the overall differences between the difficulty levels of the separate scales, and therefore by inspecting the T score profile of a group, for example, low scores provide evidence about a special weakness of the performer rather than an indication of the overall difficulty and/or of complexity of the task.

Since the score of 50 corresponds to the overall mean of the total population, the T score graphs will be presented in a form that the score 50 will represent the base line of the graph.

Graph 5.2 presents the profiles obtained in the three textbook groups.

Inspection of the profiles reveal that considerable differences between the various textbook groups appear on the fraction subtest. The "One, Two and... Three" group outscores in Fractions other groups, while in Word Problems it has a lower score than the "Simple Math" group. This finding appears to reflect differences of content emphasis in the two programs regarding Word Problems. With respect to differences in achievement on Fractions, it should be observed that does not pertain to the "Simple Math" program's list of grade four contents.

---

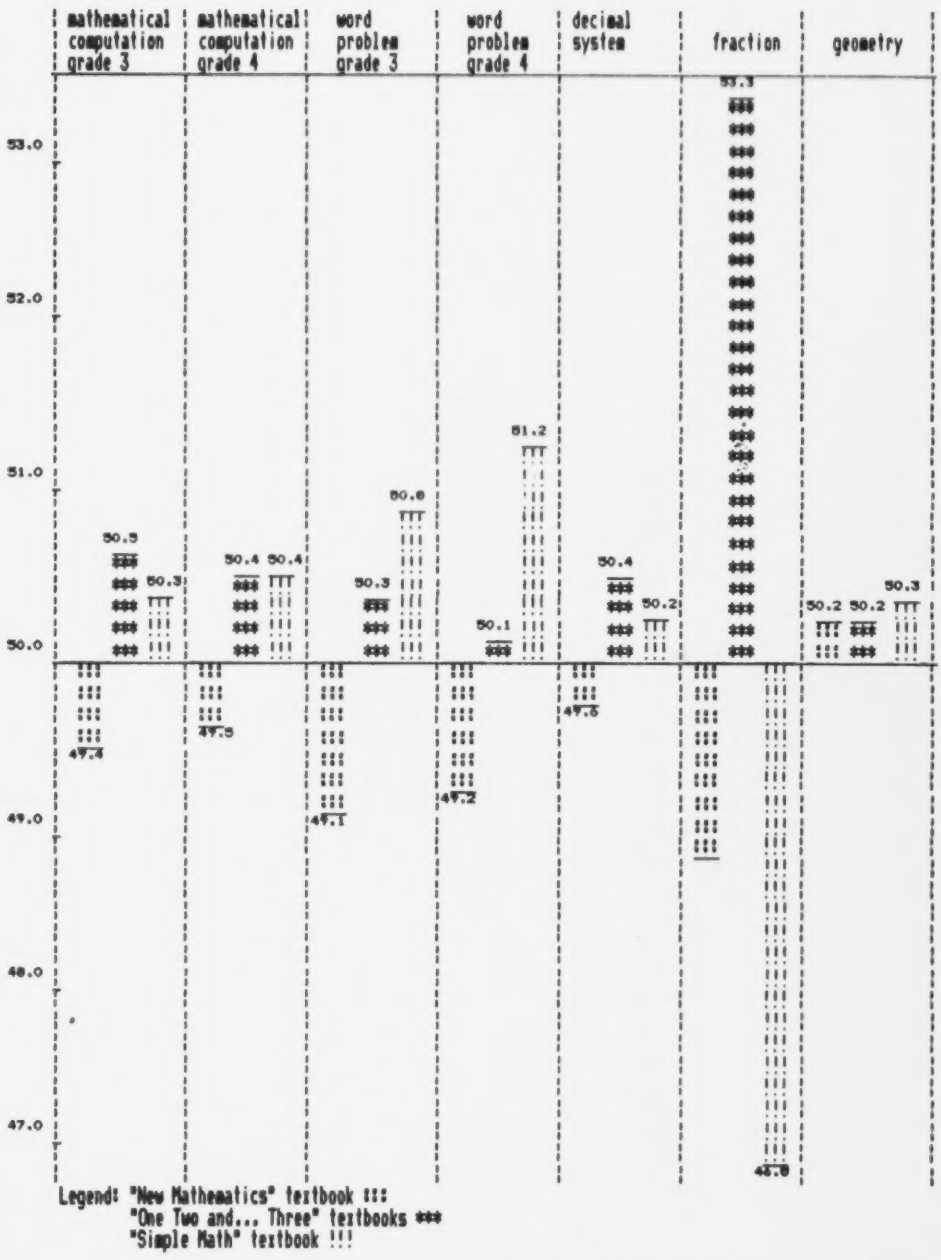
[3]The formula for calculating T score is

$$T_i = \frac{X_i - \bar{X}}{SD}$$

where: X_i is raw score;

$\bar{X}$ is the mean of X;

SD is the standard deviation.



Graph 5.2 The profiles of three textbook groups.

School Profiles

T scores were used for creating the school profiles[4] . The score on the total test provides information about the general level of achievement in a particular school. By looking at the graph one can easily read the distance of the school mean from the overall populations mean. A comparison of the subscale means with the overall mean of the school may reveal areas of relative excellence or weakness in each of the component scores.

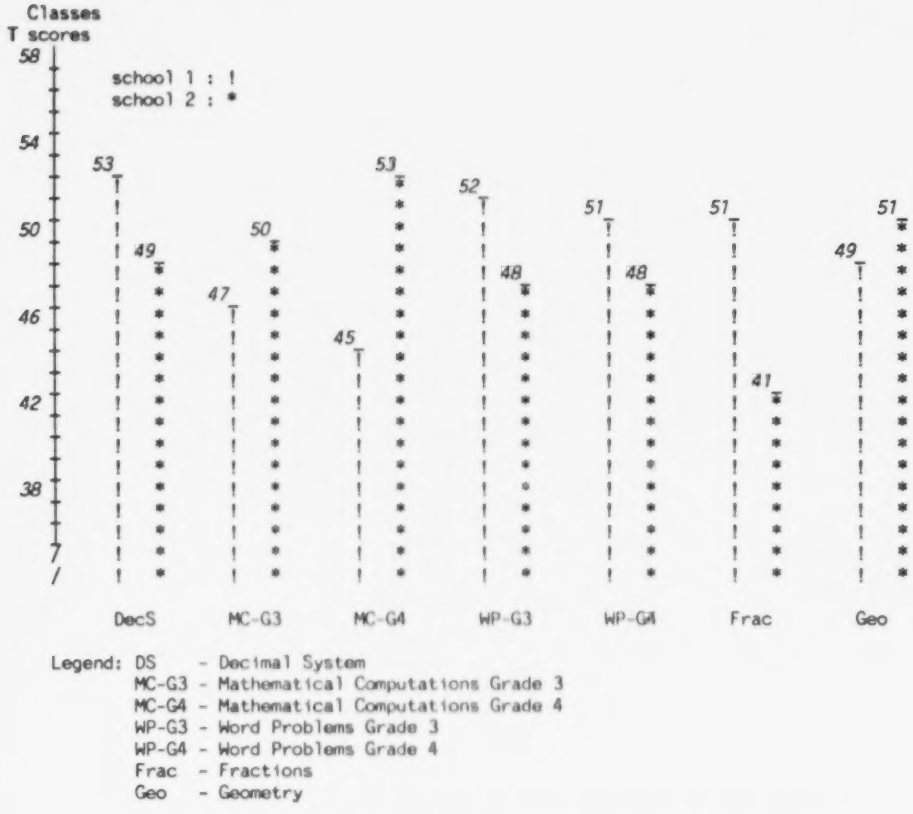
In Graph 5.3 we present the profiles of two schools which had approximately similar overall means. It can be seen, however, that they differed from each other with regard to the distribution of success, and quite likely with regard to efforts across various subareas of the test as well.

Note that on four scores (Decimal System, Word Problems Grade 3, Word Problems Grade 4, and Fractions) school 1 had higher scores and on the substests Mathematical Computation Grade 3, Mathematical Computation Grade 4, and Geometry school 2 had higher scores.

Using Nunnally's (1967) terms, the differences described above represent the "level" dimension. It may be of interest to examine differences between the two classes on the "dispersion" dimension, too. Using the simplest measure of dimension, i.e. the average distance of scores on various subtests[5], the computed values of "dispersion" for the two classes are 3.4 and 3.5 respectively. Still there are some differences in the elements which are contained in the dispersion indices. Thus, it would be useful to calculate not only the means of the dispersion elements, but also their standard deviations. In this case, the standard deviations

[4]Circumstances did not permitted the use of school profiles in actual contact made with schools. The results were transmitted to the schools in terms of percentage of scores. The discussions related to school and individual profiles present evidence on the potential use of this presentation mode of rather than the benefits derived from its actual use. It is planned that in future assessment studies, the profiles will be presented to the school together with the vectors of percentage scores.

[5]Between the seven subtests 21 such distances can be measured, such as subtest1-subtest 2, etc. All distances are expressed in absolute values, i.e. the direction of differences is disregarded.



Graph 5.3 Achievement profiles on seven subscales of two schools

are 2.2 and 2.9 respectively. Indeed in class 1 the largest distance between two scores is 8 (between DS and MC-G3) and in class 2 the largest distance is 12 (between MC-G4 and Fractions).

In comparing profiles of two groups or two individuals one may also use the correlation coefficients between the values of elements which make up the dispersion index.

The correlation coefficient of the 21 elements included (or averaged) in the discrepancy measure is .25, indicating that there is a relatively very low measure of congruence between the two

profiles.

The school profiles may serve as a basis for decision making at the school level. Changes in the curriculum structure and in instructional strategies may be recommended as a basis of analytical examination of the profiles.

Individual Profiles

The individual profiles also produced on the basis of T scores serve the purpose of diagnosis.

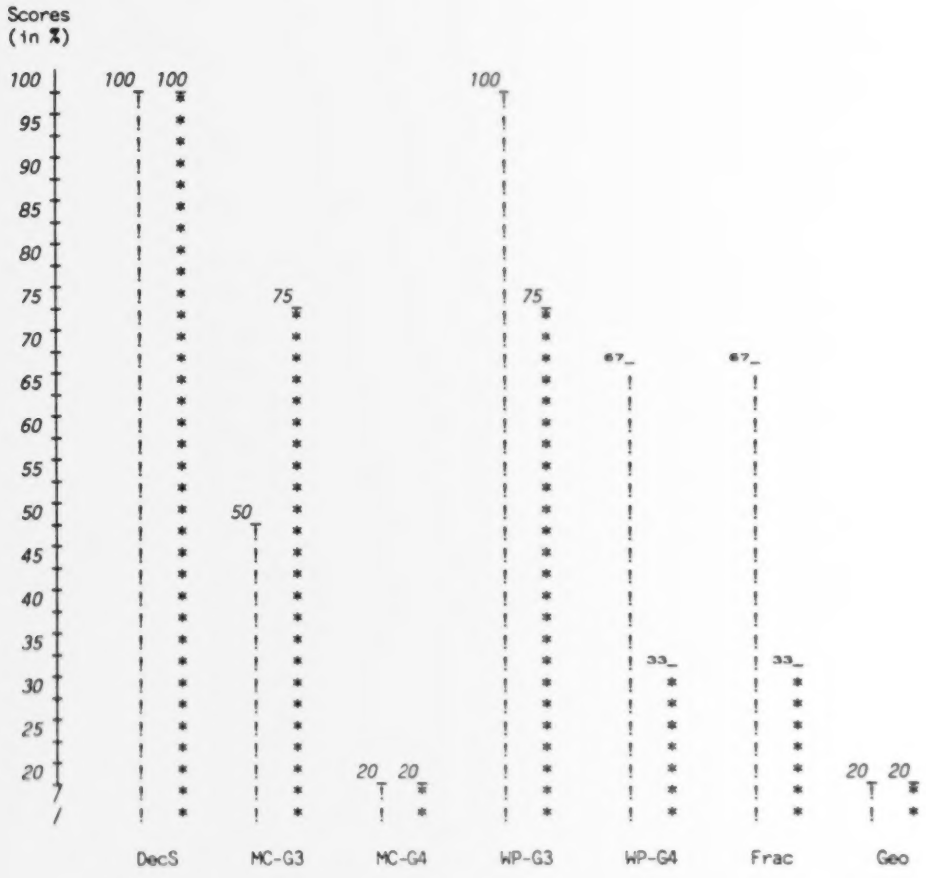
Assuming that the total score is an indicator of the overall attainment level of the individual in mathematics one may consider the subscores which are considerably different from the total score as indicators of over or under-achievement in a particular area.

A more refined diagnostic tool may be the T score profile of an individual expressed in the form of deviation from the class mean.

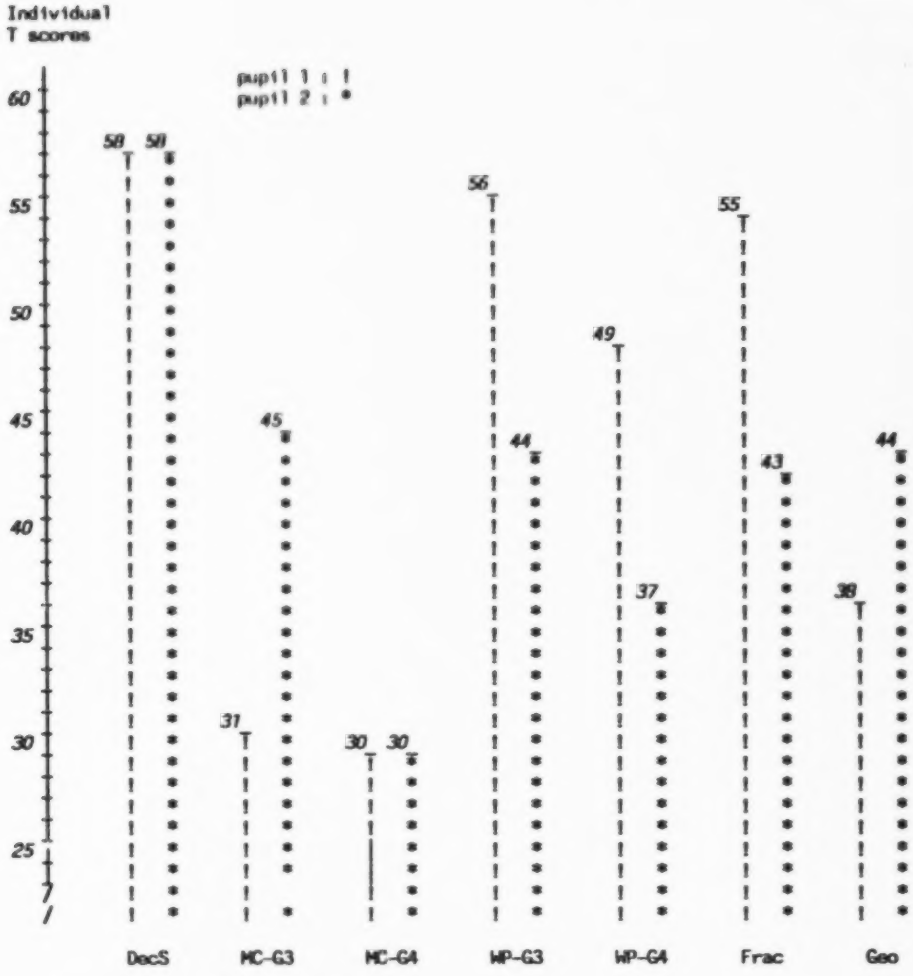
It may be instructive to observe three different modes of presenting the profile of an individual: (1) percentage scores; (2) T scores calculated based on the population mean; (3) T scores calculated based on the class means. A threefold profile of two pupils is presented across the Graphs 5.4, 5.5 and 5.6.

Looking at these three profile representation the teacher will see learn an individual learner's unique configuration of his/her achievement pattern. The threefold profile of two pupils illustrates the kind of information which can be derived from data of this kind.

The profiles appearing in Graphs 5.4-5.6 reveal, for example, that both pupils' scores on Word Problems Grade 4 are unider the national average, while on Word Problems Grade 3 pupil 1 exceeds the natonal average and pupil 2 is below it. The profiles related to the class averages are not substantially different than those related to the national average. The similarity of these two profiles is incidental, and it suggests that the class profile was quite similar to the national profile.

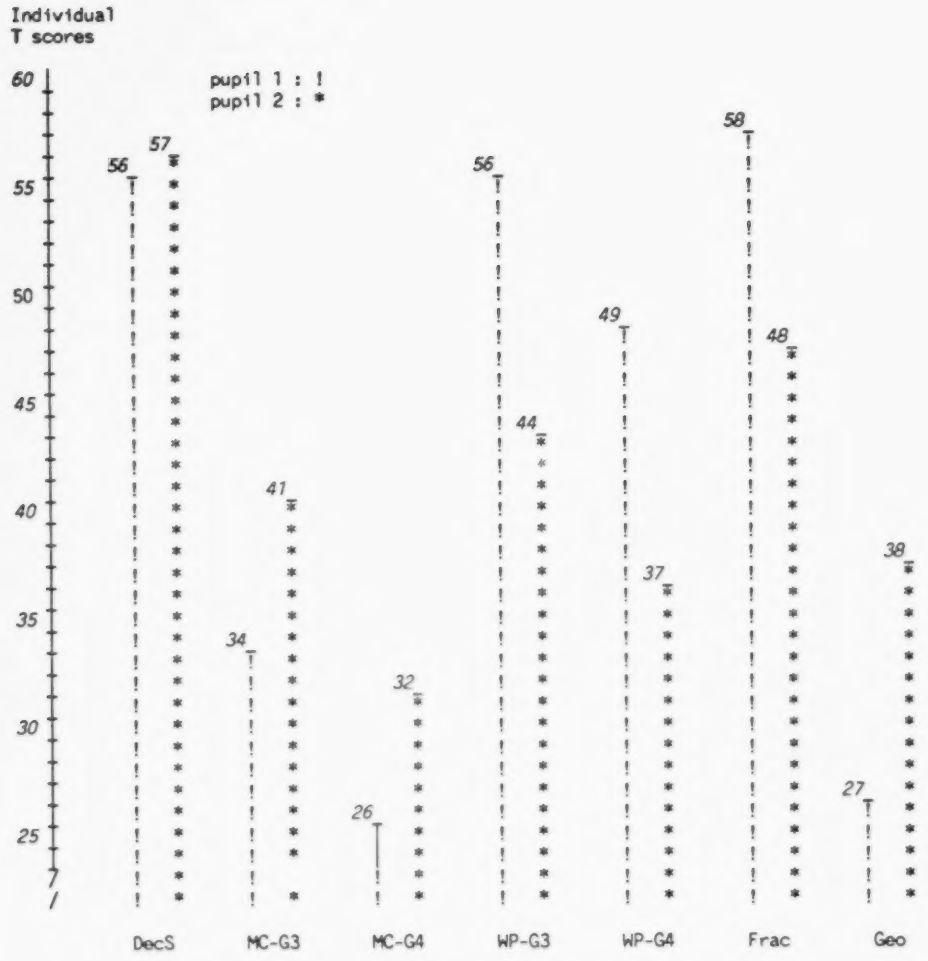


Graph 5.4 Achievement profiles (scores in percentage) seven subscales of two pupils



Legend: DS - Decimal System
MC-G3 - Mathematical Computations Grade 3
MC-G4 - Mathematical Computations Grade 4
WP-G3 - Word Problems Grade 3
WP-G4 - Word Problems Grade 4
Frac - Fractions
Geo - Geometry

Graph 5.5 Two pupils' T Achievement score profiles calculted in the form of deviations from the population means on seven subscales



Legend: DS - Decimal System
MC-G3 - Mathematical Computations Grade 3
MC-G4 - Mathematical Computations Grade 4
WP-G3 - Word Problems Grade 3
WP-G4 - Word Problems Grade 4
Frac - Fractions
Geo - Geometry

Graph 5.6 Two pupils' T Achievement score profiles calculated in the form of deviation from the classe means on seven subscales of two

Attitude Toward Mathematics

Attitude is defined as a learned disposition to respond positively or negatively to certain objects, situations, institutions, or persons (Aiken, 1985). The acquisition of a positive attitude toward mathematics is one of the major objectives of mathematics teaching. Bloom (1976) emphasized the instrumental value of a positive attitude toward cognitive achievements in mathematics. Those who like mathematics, devote more attention to it and acquire a higher level of mastery. In summarizing the first IEA's International Comparative Study of Mathematics Achievement, Husen (1967) reported that in all countries and at all grade levels mathematics achievement is correlated with interest in mathematics. Although studies dealing with attitudes toward mathematics do not provide evidence about the direction of causation, it is still a commonly-accepted view that the relationship between attitude and achievement in the subject is dynamic and interactive (Aiken, 1985).

Mathematics educators have observed that a considerable proportion of students develop a phobia toward mathematics and consequently become handicapped in their studies. To cope with this problem, several psychologists in the U.S. started to conduct mathematics-anxiety reduction seminars. According to Tobias and Donady (1977) success in dealing with mathematics anxiety begins with a public acknowledgement of of its existence. The underlying philosophy of this approach is similar to the underlying philosophy of Alcoholics Anonymous. Empirical studies have shown that negative attitudes toward mathematics develop mainly at the age of 11-13 (Aiken, 1985). Pupils' attitudes toward mathematics in Israeli schools have been examined in several studies. In a large-scale cross-sectional study Lewy and Chalfon (1973) found that on the Liking Mathematics scale very high scores were obtained across all grade levels, and at the grade 4 level only Physical Education is related higher than mathematics on the Liking-scale.

In the present study, attitude toward mathematics was measured by including nine questions dealing with with it in the Student's Questionnaire (SQ).

Responses to these questions were obtained from 2843 fourth grade pupils. The nine questions and the distribution of responses to these questions are listed in Table 5.7. The wording of the response alternatives was varied from one question to another, and it was adjusted to the nature of the question.

Therefore, instead of the actual responses only their scale code was listed in Table 5.7.

Table 5.7 Attitude toward mathematics (percentages).

Serial number	The question	1 (low)	2	3	4 (high)	All
23	Do you like mathematics?	2	2	36	42	100
29	Is studying mathematics from your textbook easy for you?	1	1	43	54	100
30	Are the topics included in your mathematics textbook important ones?	1	30	69	X	100
31	Would you like to increase the study periods for mathematics?	27	73	X	X	100
32	Is studying mathematics useful for your future?	1	99	X	X	100
33	Are mathematics studies more difficult for you than studies in other subjects?	3	40	57	X	100
34	If a mathematics lesson is cancelled are you disappointed?	44	45	10	X	100
35	Would you like to go to a career in which you need mathematics?	9	51	40	X	100
39	If instead of the mathematics lesson the teacher decide to run a Bible lesson are you sad?	47	33	20	X	100

Note: In some questions there were only three or two response alternatives. Celes marked with X represent non-existent response alternatives.

The reliability of the whole scale (using the coefficent alpha) is .720. Additionally, the scores obtained in the scale were divided into four equal intervals, the distribution of which is presented in Table 5.8.

One cannot easily define a cutting-off line on the scale which distinguishes between pupils of high and low-positive attitudes toward mathematics. Nevertheless, inspection of the frequency distribution suggests that a large majority of the pupils have a

Table 5.8 Distribution of the attitude scores.

Level	Percent of respondents
high	18
favorable	62
less favorable	18
low	2
All	100

positive attitude toward mathematics.

To complement the information contained in the Tables 5.7 and 5.8, the responses given to a single question dealing with mathematics phobia are presented here.

A straitforward question was posed to the children: "Are you afraid of mathematics lessons?"

The alternative responses and the percentage of students marking them were:
1. Very much - 1 percent.
2. A little - 13 percent.
3. Not at all - 86 percent.

Combining the responses related to mathemathics phobia with those obtained on the attitude scale, it looks as if approximately 15-20 percent of the fourth-grade pupils in elementary school have already developed negative attitudes toward mathemathics which would require some kind of corrective treatment.

Chapter Six
Factors Associated with Outcomes of Mathematics Teaching

This chapter of the study presents findings on factors associated with the outcomes of mathematics teaching. Two outcome types will be considered: cognitive and affective outcomes.

We will deal with group differences related to the major variables of the study, such as the administrative districts of the educational system, gender, the size of the class and its homogeneity, the compensatory education index of the school, and program variations.

The results presented in this section will provide partial answers to some questions which were raised in this study.

Achievement in the six administrative school districts

Since the six school districts have a certain autonomy in the allocation of resources, determining certain parameters of the school curriculum and organizing in-service training activities, it was of interest to separate data summaries for each district.

Summaries about the cognitive achievement in mathematics in the six districts are presented in Table 6.1.

As demonstrated in Table 6.1, there are substantial differences in the achievement levels observed in various administrative districts. The standard deviations in North and South are slightly larger than in other districts. Relatively large differences among the mean scores are observed on the grade four curriculum materials subtest (the range is 53-67) while on the Minimum Competency subtest and on the grade three curriculum materials subtest the range is slightly lower. Nevertheless, it should be noted that the

Table 6.1 Cognitiwe achievements in six school districts.

District		Tests						
		Total	Minimum Compet.	Calcu-lation	Decimal structure	Verbal problems	Grade 4 curric.	Grade 3 curric.
Jerusalem N=155	Mean	70	90	76	85	81	58	80
	SD	16	13	19	18	21	19	16
North N=196	Mean	65	84	70	74	73	53	75
	SD	19	19	23	24	24	22	19
Haifa N=351	Mean	77	93	84	85	83	67	85
	SD	14	10	17	18	19	18	12
Center N=933	Mean	75	91	81	83	81	66	83
	SD	17	13	19	21	21	20	15
Tel-Aviv N=795	Mean	77	92	84	85	83	67	85
	SD	16	12	18	19	19	21	15
South N=155	Mean	74	90	82	82	79	65	82
	SD	19	15	20	22	24	22	17

differences among the adminstrative districts reflect variations in input (i.e. socio-economic affiliation) rather than variation in the quality of teaching.

Gender and Achievement in Mathematics

Studies About the Relationship Between Gender and Mathematics Achievement.

Differences between the achievement levels of boys and girls in various school subjects have been of great interest to educators for the last one hundred years, and numerous empirical studies on this topic have been carried out (Keeves, 1985b). Tyler (1956) reviewed studies carried out in the U.S. on sex differences in scholastic achievements and reported that according to the findings in most cases girls obtain relatively high scores in language studies and boys obtain higher scores in mathematics. Similar results were reported by Walker (1976), who summarized the

results obtained in a cross national comparative study by the International Association for the Evaluation of Educational Achievement (IEA). Recent studies have demonstrated that the overall pattern of sex differences in mathematics achievements mask some more refined distinctions related to the outcomes of various types. Thus, Moos (1982) and Fennema (1975) reported that girls attain higher level competency in computational skills, and boys attain higher level skills in complex, and higher mental-function, types of mathematical problem solving. They also observed that the gap of the achievement level between boys and girls in mathemathics increases in secondary school compared to primary school.

Researchers suggested three different explanations for the sex differences in mathematics achievement: role perception; personality traits; and, differential interaction of environment and personality (Schildkampf-Kundiger, 1982).

(1) Role perception - Women rarely choose careers which a require high level of knowledge in mathematics. In the U.S. only 3% of the engineers, 18% of life and physical science specialists, and 23% of computer specialists are women (Bureau of the Census, 1980).

In Israel 40% of the recipients of a B.A. mathematics and 30% of the recipients of higher degrees in mathematics (M.A. and Ph.D.) are women.

It is of interest to note that till the middle of the 19th century, mathematics studies were not considered to be a domain of masculine interest. Pearl (1979) analyzed the 1704-1841 volumes of the popular Ladies Diary and found no trace of sex stereotypes related to mathematics. However, the changing social role of women in the 19th century alienated them from studying mathematics, and only recently, as a result of the women's liberation movement, one can observe changes in the careers that women choose.

(2) Personality trait - In 1873 Herbert Spencer, in an article entitled "Psychology of the sexes", claimed that women developed differently in the course of evolution (Keeves, 1985b). One of these differences is the woman's inferiority in mathematics ability. Analyses of measured patterns on intelligence tests factors have been frequently brought up in support of this theory. E. L. Thorndike in 1906 rejected this view, however, since, according to his opinion, differences are so small that they do not have prac-

tical significance (Sherman, 1977).

(3) Interaction between environment and personality - Schildkampf-Kundiger (1982) offers a cognitive view of the achievement motive to explain the sex-related nature of mathematics achievement. She claims that in mathematics lessons pupils have to solve intellectual problems, and that young girls facing such achievement situations are confronted with conflicting behavior standards. On one hand, they have to consider the school's achievement norms, on the other they have to act in a way which fits their sex-roles stereotype. Indeed, she found that girls having lower achievements in mathematics than predicted (on basis of their measured intelligence levels --underachievers), perceive mathematics as a masculine study area, whereas the overachievers believe that mathematics study is suitable for women, too. Boys perceive the utility of studying mathematics differently than girls, and consequently they react differently when confronted with difficulties in mathematics study.

Differences between the achievements in mathematics of boys and girls have been studied in surveys in Israel, too. Lewy and Chen (1977) found that in grades four and six the difference in mathematics achievements, in favor of boys, was around 10-15 percent of a standard deviation. Minkovits et al (1977) calculated the within class partial correlation coefficients between sex and mathematics achievements (boys coded as 1, and girls coded as 2) and, after controlling for classroom context variables, found correlation coefficients approaching zero. It is interesting that in a study of sex differences in the Kibbutz, Dar (1974) found a correlation of .40 between sex (boys scoring higher than girls) and matriculation examination achievements in mathematics. Dar explains this supprisingly high level of correlation as an effect of the division of men and women in the kibbutz. According to Dar despite the egalitarian ideology of the kibbutz, most women in the kibbutz work in routine services and, therefore, girls have little motivation to study mathematics.

Sex Related Findings in the Present Study

The differences between boys and girls in mathematics in the present study were examined for the total population and for all 99 schools of the sample.

The correlation between sex (boys coded as 1, and girls as 2)

and mathematics achievement is -.07.

Univariate statistics about the achievements of boys and girls are presented in Table 6.2.

Table 6.2 Mathematics achievement levels of boys and girls

The test	Number of items	Boys (N = 1436)		Girls (N = 1407)		Difference in units of SD.
		Mean	Std.Dev.	Mean	Std.Dev.	
Total Test	40	75.2	17.0	73.8	17.4	.08
Minimum Competency	10	91.3	13.2	90.6	13.9	.05
Grade 3	21	83.8	15.4	82.3	16.2	.09
Grade 4	19	65.7	21.3	64.4	21.5	.06
Mathematical Computation	13	81.9	19.3	81.0	20.0	.05
Mathematical Computation Grade 3	8	85.5	18.3	84.6	19.0	.05
Mathematical Computation Grade 4	5	76.2	27.9	75.1	28.3	.04
Word Problems	7	81.8	21.4	79.9	21.7	.09
Word Problems Grade 3	4	88.4	20.8	86.9	21.8	.07
Word Problems Grade 4	3	73.1	30.2	70.6	30.3	.08
Decimal System	3	83.1	20.5	82.9	21.0	.01
Geometry	3	54.6	30.5	52.1	30.1	.08
Fractions	5	57.1	30.6	57.7	30.1	-.02
Measurement	2	71.9	34.0	69.1	34.2	.08
Prime and Decomposable Numbers	2	64.7	41.0	61.5	42.2	.08
Approximation, Divisibility signs, Basic Rules	5	74.7	23.8	72.7	25.1	.08

Additionally, differences in mathematics achievement between boys and girls were calculated for each school separately. We present in Table 6.3 descriptive information about differences in

the proportion of girls in the classes.

Table 6.3 Distribution of girls in the classes.

Percentage of girls	Number of schools
less than 20	2
21 - 40	12
41 - 60	73
61 - 80	8
81 - 100	2
Total	97

The differences in achievements of boys and girls on the total test are presented in Table 6.4

Table 6.4 Distribution of achievement differences of boys and girls on the total test.

The difference in in favor of boys	Number of schools
less than -13	1
-12 to - 8	6
- 7 to - 3	17
- 2 to 2	32
3 to 7	19
8 to 12	12
13 to higher	7
Total	94

It can be seen that in 32 classes the difference approaches zero. In 38 classes boys had a higher average and in 24 classes girls had a higher average score.

It was also of interest to examine the relationship between the proportion of the girls in a class and the achievement differences in mathemathics. In the IEA studies it was found that, in several countries, girls in coeducational schools had higher level achievements in mathematics than in schools for girls only. The researchers interpreted this finding as an indicator that the school context affects the sex orientation of mathematics studies. Thus, we were curious to see whether or not the proportion of girls in a class in the fourth grade of Israeli school is associated with sex differences in achievements.

In order to examine the exact nature of the relationship between the sex composition of the class and sex-related achievement differences we present detailed information in Table 6.5. Summarizing the findings of Table 6.5 one may say that no relationship was found between the proportion of girls in the classes and their class-mean scores.

Table 6.5 The relationship between sex composition of classes and sex-related achievement differences over classes.

Achievement differences in favor of boys	Percentage of girls in classes				Total number of classes
	less than 20%	21 - 40	41 - 60	61 - 80	
less than -13			1		1
-12 to - 8			5	1	6
- 7 to - 3		4	11	2	17
- 2 to 2		2	27	3	32
3 to 7		5	13	1	19
8 to 12			12		12
higher than 13	1	1	4	1	7
Total number of classes	1	12	73	8	94

Schools with Different Social Priority Rating[1]

As indicated by Lewy and Chen (1977),

> *One of the leading challenges presenting itself to educational systems all over the world today is reducing the gaps of educational achievement levels reached by people from various social strata... The existence of an educational achievement gap is a severe problem particularly in societies in which belonging to the lower class often coincides with membership of a particular ethnic group, as is the case in United States and to a large degree also in Israel.* (p.5)

During the 40 years of its existence leaders of the educational system in Israel, as well as the whole Israeli society, have been greatly concerned with the achievement gaps measured between children of oriental origin (Islamic countries of the Middle East) and children of European and American immigrants.

Kleinberger (1969) observed that immigrants of oriental origin came from traditional, semi-feudal, underdeveloped countries and were quite unprepared for the demands placed upon them by a modern industrializing society. Over these forty years the Israeli educational system installed a great variety of compensatory and fostering programs for the children of oriental immigrants and, according to achievement surveys, the achievement gap between the two major ethnic groups has been considerably reduced. While (according to early achievement surveys, based on data up to 1970) the magnitude of the achievement gap reached the level of approximately one standard deviation (Lewy and Chen, 1977), the gap has been reduced to about 80 percent of one standard deviation. Even though there are indications that the achievement gap was further reduced in the 1980's, there is still an on-going interest in the achievement differences of ethnic groups in Israel.

Due to governmental policy, the school records available to researchers do not contain information about the ethnic membership of an individual learner. Therefore, instead of dealing with

---

[1]The terms "Social Priority Rating" and Compensatory Education Budget Index" are used in this report interchangeably.

ethnic achievement differences directly, we will use a proxy measure. The ministry of Education and Culture allocates special funds to cover expenses of compensatory education programs for the disadvantaged. This budget is distributed among schools on the basis of the proportion of children from disadvantaged social strata studying in a particular school. To facilitate the distribution of funds across schools each year, the Ministry publishes a social priority rating index for each individual school, which serves as a basis for determining the magnitude of compensatory funds a school is entitled to receive. While this index is not fully related to the proportion of oriental children in a particular school it may serve as an acceptable proxy for describing the ethnic composition of the school. Since in the present study there was no possibility of determining the ethnic membership of the individual learner, the ethnic differences were estimated on a basis of contrasting achievements in schools which had a low-level priority index (i.e., assumedly did not have, or had very little proportion of children from disadvantaged social strata) versus schools which had a high rating index (i.e., assumedly had a high proportion of children from disadvantaged social strata). It should be emphasized again that this measure constitutes a proxy for the direct measure of ethnic differences, and sufficient information is not available for determining the adequacy of this measure. Therefore we exercised caution, and referred to findings reported in this section as achievement differences between learners studying in schools of low and high social priority indexes. This may contain hints concerning ethnic differences, but by no means do we claim that these findings constitute a reliable basis for making inferences about ethnic differences.

The priority index of the Ministry of Education and Culture is expressed as a scale having values of 0 to 100. The indices were clustered into three groups: high achievement group (i.e. low priority index: 0-13 index values); middle level (index value 14-36); low achievement group (high priority index: 37-100 index values).

The results of the three priority group pupils on the test are presented in Table 6.6.

It can be seen that the differences between the three groups are low. Relatively large differences are observed on the "word

problems". It may well be that pupils in the low achievement group have difficulties in reading comprehension and therefore they failed to deal with "word problems".

Table 6.6 Cognitive achievements of three priority groups

Priority Index		Tests						
		Total	Minimum Compet.	Calcu-lation	Decimal structure	Word problems	Grade 4 curric.	Grade 3 curric.
High achievement N=1172	Mean	76	91	81	84	83	66	84
	SD	17	13	19	20	20	21	15
Middle achievement N=1014	Mean	75	91	83	84	81	66	84
	SD	16	12	18	20	20	21	15
Low achievement N=611	Mean	72	89	80	81	76	62	80
	SD	18	16	21	22	24	22	17

Users of Various Programs

The scores on the mathematics test were separately summarized for groups of schools using the program. Table 6.7 contains relevant information for schools that used the three most popular programs.

No substantial differences were observed between the mean scores of various groups. The range of the mean scores on the total test is 73-76 and on the Minimum Competency test 90-92. Greater differences were oserved on the "word problem" items (the range is 77-84).

Table 6.7 Cognitive achievements in three program groups

Instrutional program		Tests						
		Total	Minimum Compet.	Calcu-lation	Decimal structure	Word problems	Grade 4 curric.	Grade 3 curric.
New Mathematics N=538	Mean	73	90	80	82	77	62	81
	SD	19	15	21	20	24	22	17
Simple Math N=586	Mean	74	92	82	83	84	64	83
	SD	15	12	18	20	19	19	14
One, Two, and Three.. N=929	Mean	76	91	81	83	81	67	83
	SD	16	13	19	22	21	21	15

Chapter Seven
The Utilization of Evaluation Results

For a very long time the decisive criterion, if not the only one, for evaluating the merit of an evaluation study was its methodological adequacy. In the 1970's one witnesses the emergence of an additional criterion: its utilization. In a seminal paper Weiss (1972) claimed that the primary justification of evaluation is utilization. Alkin et al (1979) conceived the issue of utilization as the distinguishing feature between research and evaluation, claiming that research accumulates knowledge, while evaluation purports to serve the information needs of decision makers. Decisions related to curriculum development and implementation which may be affected by evaluation results are listed by Davis (1981). They are: modification of the instructional materials, selection of content elements, qualifying the utilization of the program, and adopting or rejecting the program. The list of instances which represent utilization of evaluation results has been substantially expanded in the recent research literature. Alkin et al (1979) distinguished between the restrictive, mainstream definition of evaluation utilization and its extended, alternative definition. The restrictive definition refers to decisions relying solely on the evaluation results and made immediately after their publication by those who commissioned the study. In this restrictive definition, reference is made to four specific issues: time (immediately after the publication); users (the authorized recipients); the topic of decision (the reason to commission the study, i.e. alternative ways of action explicitly defined in the study plan); the input used for decision (only the reported results). The alternative definition substitutes general for specific terms in the sentence describing utilization. It then expands its scope to action taken at anytime, by any authorized or interested person, with regard to a variety of problem areas related to the target of evaluation, either solely on the basis of the evaluation input or on the evaluation input in combination with informations derived from sources.

Moreover, Alkin and Daillak (1985) attributed utility to the impact of evaluation studies --which results in no decision making at all-- such as changing attitude, creating dispositions, increasing certainty, raising doubts, sparking consideration of a certain issue, focusing attention on a problem, etc.

It may be of interest to list here a few historical instances of evaluation utilization in Israel, which may serve as a background to an examination of patterns of use of evaluation results adopted in the present study. It may also advice on possibilities for further utilization of the present study, and of the other series of assessment studies, which have been carried out within the framework of the Israel Curriculum Center.

The first instance of the utilization of evaluation studies in Israel was the 8th grade SEKER (achievement survey) examination.

For a period of two decades (up to 1973) the annual SEKER examination was administered to all 8th grade pupils in the Israeli educational system (50,000 - 60,000 pupils at various periods during that time). The examinations served as a screening device for admission to the academic-type secondary school (Ortar, 1967). In this case, the country witnessed direct utilization of evaluation results in a way which was compatible with the mainstream definition of the term. But the utilization of the SEKER examination exceeded the boundaries of the mainstream type utilization. Its national importance and lasting impact on the Israeli educational system has been due to one of its unintended side effects: it provided systematic information about the achievement level of children from various socio-economic strata and focused attention on the achievement gap between children whose parents immigrated from industrialized (European-American) countries and those whose parents immigrated from developing (Asian and North Africa) countries. The results of the SEKER examination gave a great impetus to the installation of compensatory educational programs for learners from disadvantaged social strata.

A second example of utilizing evaluation results is a series of evaluation studies conducted by the Israel Curriculum Center with the aim of evaluating school curricula and instructional materials. The mainstream type of utilization of formative evaluation studies conducted within this framework is illustrated by Lewy (1973). The indirect and expanded scope in the utilization of evaluation studies for improving the curriculum material quality is described in detail in a recent publication of the Israel

Curriculum Center (Lewy, 1986).

A third example of utilizing evaluation results is a series of nationwide grade-three, basic skills assessment studies. In 1983 and onwards, the Chief Scientist of the Ministry of Education and Culture commissioned the Szold Institute to test basic literacy and numeracy of pupils in grade three of elementary school. The purpose of these testing projects was to identify children who are in need of corrective teaching, and to help them master basic literacy and numeracy skills (Razel et al, 1984). The results of these assessment studies enabled officials to prescribe corrective teaching for those who failed the tests. And, as reported by the Chief Scientist, in the wake of these studies innovative, corrective teaching methods in reading comprehension were developed, schools payed more attention to teaching reading skills, and the reading comprehension level of third-grade elementary school children rose 10 percent (R. Raz, personal communication). Again, this is an example of research utilization compatible with both the mainstream definition, and the expanded, alternative definition of the term.

Finally, as an additional example of evaluation utilization, one may mention the Van Leer Assessment Study (Minkovits et al, 1977). In this study, the conditions of teaching, and the educational achievements of first, second, fourth and sixth-grades in elementary school were examined in a sample of 98 schools.

The purpose of the study was to provide a comprehensive picture about three aspects of the elementary school in Israel:

1. resources and their allocation to schools of various types;
2. achievements in major subject areas (reading, arithmetic, biblical studies, geography, and science).
3. home and school variables associated with achievement level.

This study had an unusually great impact, and its results were highly publicized throughout the country. The media, professional organizations, teachers' in-service training courses, all devoted great attention to the results and conclusions of this study. The Pedagogical Secretariat of the Ministry of Education jointly with the Israel Curriculum Center established several committees and imposed upon them the task of formulating specific recommendations based on the study's findings. The impact of the study was examined and described in detail by Lewy and Alkin (1983), who also

tried to identify the reasons for its impact in Israel.

The four examples described above illustrate that the overall situation related to research and evaluation utilization is not to be viewed as been so pessimistic in Israel as it is in the United States. In Israel the notion of the "graveyard of disregarded evaluation studies" is not as current as elsewhere. And "gloomy statements about the impotence and futility of evaluation" are less common in Israel than in the U.S. (see Alkin and Daillak, 1979, p.13-14). Boruch (1981) attributed these differences between the United States and Israel to the fact that the Ministry of Education and Culture in Israel made institutional arrangements to promote research and evaluation utilization by creating the office of the Chief Scientist. The office of the Chief Scientist exists in several Ministries in Israel (Agriculture, Transport, Commerce and Industry, Health, etc.). Its head is a renowned scientist, usually a well known university professor from a relevant scientific discipline. His/her task is both to supervise research activities commissioned by the Ministry, and to monitor the utilization of the results. Thus the sponsor's direct involvement in, and control of a study is not terminated with the delivery of the study report, as often happens in the U.S., but the Chief Scientist is entrusted with the task of looking after the utilization of the results.

These favorable conditions do not guarantee the optimal utilization of research or evaluation results. Indeed, one could also list a series of evaluation studies which had little impact in Israel. As indicated by Patton (1985), it is a part of the evaluators' responsibility to create an interactional pattern with the evaluation clientele to facilitate the utilization of the results.

The present study was designed with the purpose of serving as a basis for decision making about, and an improving of, mathematics teaching at the lower grades of elementary school. It was expected that some decisions would be made immediately after publishing partial, interim data summaries, and that the results of the study will have effect forthcoming years. This being the first comprehensive report of the study, it may well be premature to assess its overall impact and to specify what scope of decisions will be taken as a results the findings.

Therefore, we will deal only with selected aspects of utilization: (1) Features of the study which facilitated the utili-

zation of the results; (2) Expected and actual decisions taken on the basis of the study results; and (3) Suggested ways to increase utilization. Finally, further recommendations for action will be presented in the concluding chapter of the study.

Features of the Study which Facilitated Utilization

The study was conducted with the aim that the results should be utilized for decisions at the following levels: the individual pupils, the school, the administrative school district, and the national system of education. Patton (1985) claimed that placing priority on the utility of evaluation has profound implications for every aspect of the evaluation process. He presented a list of premises, which are related to utilization, and these premises serve as a framework for describing aspects of the present study which facilitated utilization.

Planning for Utilization

Concern for utilization is an ongoing process from the very beginning of the study. "*Utilization is not something one becomes interested in at the end of the evaluation.*" (Patton, 1985, p. 1748). After completing an evaluation study, the potential of its utilization has largely been determined. Keeping this in mind the researchers established a Steering Committee for the study. The members of the Steering Committee were appointed by the Director of the Israel Curriculum Center and they represented various interested "stakeholders" of mathematics education in the elementary school: teachers, supervisors, and mathematics curriculum specialists. In the Israeli educational system, all activities related to teaching a particular subject (including curriculum development, approval of textbooks, in-service training, supervision of pre-service teacher training) are coordinated by a Subject Superintendent. The Subject Superintendent for Mathematics was the chairman of the Committee. The Committee was convened at various phases of planning the study. They reviewed the drafts of data-collecting instruments and scrutinized the utility of each bit of information to be collected. During the phase of data analysis, the Committee specified the kind of data summaries which they judged to be useful. The Committee also took active part in disseminating the findings of the study. Neither the Committee nor the study team has been dissolved after the publication of the

first evaluation report. Though there are no more formal Committee meetings, the Committee members are still entitled to request further data summaries from the evaluation team, and such requests are fully respected. Indeed, the continuous work of the evaluation team, in an institutional framework, engaged in conducting a series of evaluation studies, greatly facilitates the extended utilization of both the evaluation findings, and the available data base for producing additional data summaries as requested by those who need information.

Oritentation Toward the User

Another premise listed by Patton (1985) is that evaluation should be user-oriented. "*This means that the evaluation is aimed at the interest and information needs of specific identifiable people, not vague passive audiences.*" (Patton, 1985, p. 1748). Therefore the evaluators have to indentify the potential users of the information.

The importance of identifying the potential users is emphasized also in the Joint Committee's (1981) list of Utility Standards.

It suggests that

> *Audiences involved in, or affected by the evaluation should be identified so that their needs can be addressed* (p. 21).

The authors of the Standards suggest that the evaluation audiences (or clienteles) should be rank ordered, and within the resource and time limitations, the evaluators should strive to accomodate their information needs.

The Steering Committee, entrusted with the task of determining the major parameters of the study, consisted of representatives of various groups of clientele, or stakeholders, identified by the evaluation team. During the Committee meetings the evaluators tried to learn what expectations the various adiences had from the study, and what kind of information they would be willing to use.

Users' Personal Involvement

The interested decision makers and information seekers should

be personally involved in planning the study, monitoring the whole process, and supporting the dissemination of the results.

> *Working actively with people who are stakes in the outcomes of an evaluation (the "stakeholders") is aimed at increasing the potential for utilization by building a genuine commitment to and understanding of the evaluation over the course of the evaluation process* (Patton, 1985, p. 1748).

In the present study the stakeholders were personally involved in the development of the data collecting instruments and in determining the form of the data summary. They participated in discussions related to the interpretation of the results, and they were asked to express their opinion about the preliminary version of the recommendation. The stakeholders derived satisfaction from their involvement in the study, and they gained reputation among collegues for participating in such activities. They invested time and effort to improve the quality of the evaluation and were eager to obtain information about each step of the study.

Multiple Clientele

The Committee consisted of representatives of various interest groups: teachers, school supervisors, mathematics educators from the Ministry of Education and from the academic world. Curriculum developers also participated in its work. The evaluation team strived to develop sensitivity to, and respect for, the varied and multiple interests.

To respond to the information needs of various "stakeholders", shortly after completing the data collection the evaluation team produced two different written reports and conducted several workshops. The present study constitutes the third written report illuminating selected aspects of teaching mathematics in the Israeli elementary school. An abbreviated resume of this report will be published (in Hebrew) by the Israel Curriculum Center.

The content of the reports, their presentation style and format, and also the timing of their publication were all geared toward satisfying various groups of clientele. Moreover, even the timing of the data collection and the priority order of data-processing activities were determined in such a way as to best provide a balanced response to the multiple information needs.

The data were collected at the end of the second trimester of the Israeli school year. This time schedule was imposed upon the evaluation team, in spite of the fact that at the time of the testing, teachers had not covered the whole syllabus prescribed for grade four. It was necessary to limit the content of the test to topics which usually are covered in schools before the end of the second trimester. The schedule ensured that classroom teachers would receive the results of the testing before the school year was over. The evaluation team felt that it woud be of little value for teachers to receive information about their pupils' achievement after the school year was over. Many teachers may transfer to other schools, or may assume responsibility for teaching in another class. Thus it seemed to be advisable to deviate from the common pattern of conducting testing at the end of the school year. The possibility of providing quick feedback to teachers took precedence over the desire of the evaluators to cover the entire grade four mathemathics curriculum in the test.

The first report prepared by the team was designed to satisfy the information needs of teachers. Its details will be described in the next section of this chapter. The second report came served the needs of educational administrators and curriculum planners, helping them make decisions about issues which may require immediate action.

Finally, the present study, and its resume which will be published in Hebrew and disseminated to experts and professionals in mathematics education and in educational research, aim to summarize knowledge derived from the study. Accumulatively, the study may also contribute to our better understanding of issues related to teaching mathematics in Israeli schools. It was designed also with the aim of serving as a base-line data set for examining trends in mathematics teaching and achievement in the future, and as source materials for experts in the preparation of communiques to the mass media, and for teachers familiar with problems of teaching mathematics.

Expected and Actual Decisions Taken Based on the Study Results

Despite the meticulous planning of the study and the early and intensive involvement of representatives of various stakeholders, a gap was observed between the expected and actual level of uti-

lizing results. In another context, Finley and Berdie (1970) called attention to the fact of how difficult it may be for expert evaluators to estimate the actual level of a particular skill in a given target population. It looks like setting realistic expectations for utilization, and determining what is a satisfactory level of utilization, also requires great experience.

We will deal with issues of utilization separately for various users groups.

Utilization by Classroom Teachers

It was expected that teachers will use the information for the following purposes: diagnosing learning difficulties at the whole class level and at individual pupil level; planning remedial teaching activities; and improving testing procedures in their class.

It was also expected that the impact of the testing would be felt not only in the schools included in the sample, but also in other schools. It was expected that teachers from other schools would be willing to take advantage of the test, and by using it derive information about the achievement level of their pupils in comparison to national norms supplied by the evaluation team.

It should be noted that the entire test used in the study was released to teachers, and there was no restriction of any kind for free utilization of the test in all schools.

To motivate teachers to use the test results, a short workshop was held in Tel-Aviv approximately ten weeks after the testing completion, and one month before the end of the school year.

Teachers from schools included in the sample were invited to attend the workshop (100 schools) and approximately 75 schools sent representatives to the workshop.

The teachers received a handout which contained information about the univariate statistics in the entire sample of scores on the whole test (total score) and subtests; the difficulty level of each item in the whole population; and percentile norms for individual scores and school averages. Additionally, each teacher received detailed information about the pupils' scores in his/her own class; the mean achievement level of his/her classe; and the

difficulty level of each item in his/her class.

Teachers were taught how to create profiles of individual achievements and how to derive individual and class percentile means for their own class.

A half year later a limited follow up was carried out to examine the utilization of the study results at the class level. A site visit was carried out at three schools in the sample, which was then followed by a telephone interview with four more teachers. The information about utilization of results are based on interviews with seven teachers (three during the site visits and four over the telephone). Surprisingly enough, the major concern of the teachers focused on two issues: the validation of their own scores assigned to the pupilse; and the use of the results of the mathematics test for the purpose of establishing ability groups within the class.

Little concern was focused on the diagnostic information provided by the scores, and little interest was revealed in the relative status of individuals and the schools in the total population.

There were instances of using the same test in parallel grade 4 classes in the same school. There was no chance of obtaining information about the utilization of the test in schools not included in the sample.

Utilization by Regional School Supervisors

A second brief report for supervisors, mathemathics education experts, and curriculum developers was disseminated by the Israel Curriculum Center. The long summer vacation and the autumn Jewish holiday season required a delay in all activities related to the utilization of the results by supervisors and other stakeholder groups (except teachers), and thus the second report was disseminated only eight months after the data collection. The release of the second report was followed by two conferences in the Tel-Aviv area. The purpose of the meetings was to present the findings of the study, to discuss with stakeholders the implications of the study for their work, and to ensure their cooperation in forthcoming evaluation studies.

Due to the countrywide concern with problems of minimum compe-

tency in basic skills, the supervisors focused their interest on the failure rate on the minimum competency set of items. Two questions were of great concern: the number of children who failed the standards set on the minimum competency tests, and the differences among schools with regard to the "density of failures" in a single school, i.e. how many children failing on the test are concentrated in the school. The information contained in Tables 5.2 (see Page 119) and 5.4 (see Page 124) of this report was of special interest for this group. Nevertheless, there were two obstacles to the direct utilization of these findings. First, data were collected only in a sample which consisted of less than 10 percent of the total schools, and thus several supervisors were in charge of a group of schools in which none of them was included in the sample. Secondly, the policy of the evaluation team was to present school-related results only to the staff of that particular school. Such an arrangement was made in order to ensure the collaboration of the school with the study team. Both the evaluation team and the schools wished to avoid having the results used by a supervisory body, which might have held them accountable for low achieving and would then impose penalties. In this context, it is of interest to know that supervisors exerted pressure on the evaluation team to identify schools included in the sample and to reveal the results obtained in each one of these schools. It was difficult to reject this demand, but the evaluation team remained firm and did not succomb to the demand.

The expectations concerning the utilization of the results by supervisors should be divided into two categories: expectations envisaged at the phase of planning the study, and expectations which emerged after inspecting the study results. At the planning phase it was expected that supervisors will take advantage of the norms of achievements derived from the results of the study and will administer the test in schools both for determining the relative level of achievements in various schools, and to advise teachers on how to identify low-achieving students in their classes. After inspecting the results of the study, one of the salient findings was that in a substantial proportion of schools (approximately 35 percent) teachers do not participate in decisions about selecting mathematics instructional programs, and in approximately 35 percent of the schools, problems related to teaching mathematics were not discussed in teachers' conferences at any time during the school year. It was expected that supervisors would take appropriate action and motivate school principals both to increase teachers' involvement in selecting textbooks and other instructional materials, and to organize teachers' con-

ferences dedicated to discussing problems of mathemathics teaching.

No systematic follow up was carried out to ascertain whether the supervisors have indeed initiated such activities. Yet, there are some signs of positive reaction. Several supervisors asked the evaluation team's assistance in carrying out testing in schools of which they are in charge. Unfortunately, the assistance was restricted to giving them copies of the examination. The evaluation team was not able to provide computer services for scoring the tests. Nevertheless, an awareness of the possibility that achievement testing projects could be carried out on behalf of regional school supervisors, was created.

It is still early to determine whether regional school supervisors acted successfully to increase teachers' involvement in decisions within the school. The great interest which supervisors revealed in discussions with the evaluation team suggests that one may expect on-going efforts by supervisors in this direction.

Utilization by Curriculum Developers

As indicated in the previous sections, several curriculum development teams prepared instructional programs for elementary school mathematics in Israel.

Some of the programs have been developed by teams operating within universities and research institutes, and they enjoy the support of highly qualified experts. Other programs have been developed by experienced teachers working without institutional support. It is quite natural a competition exists between these various programs, and those in charge were eager to receive comparative results about the achievement level of pupils using various programs. To some extent, the Israel Curriculum Center conceived the study as fulfiling two roles: evaluating curricula used in the system, and assessing achievements in mathematics. The curriculum evaluation aspect of the study was mainly concerned with providing a portrayal of various programs and with presenting a profile about strengths and weekenesses of different programs. However, the project teams had a different conception about the meaning of evaluation and about the significance of this particular study. They evinced interest mainly in the overall test scores of pupils using their programs, and viewed the study as a means of demonstrating the superiority of the programs they developed.

Thus there was a discrepancy between the expectations of the evaluation team and the program teams concerning the utilization of the evaluation results. The evaluation team expected that the curriculum development teams would focus their interest on the implications of the evaluation results for improving the program implementation, preparing adequate supplementary materials, changing contents of in-service training programs, and harvesting ideas for revising the programs or at least some elements of the program when new editions will be produced. It seems, however, that the program teams were hoping to prove the overall superiority of their program, and were worried that the results might disappoint them. The discrepancy of expectations explained why representatives of the programs were the most-involved participants in the committee-work, and why they took positions on a broad scope of design issues. At the planning stage, great interest was evinced in selecting items for the test, and representatives of the curriculum programs vetoed several items proposed by the evaluation team. During the phase of data analysis the preliminary results were communicated to those projects who revealed insterest in them. Members of some programs meticulously studied the results and carefully noted all data which provided favorable information about the effectiveness of their own programs.

The report published by the Israel Curriculum Center contained comparative information on the conditions of teaching, as well as on pupils' achievement, with special regard to the users of the three most widespread instructional programs in the system. It clearly stated that none of the programs was found to be superior, and consequently a recommendation was conveyed to the "consumer community" to continue to use programs which they prefer. These results and recommendations encouraged the curriculum program teams to shift their attention from the comparative aspects of results, to those findings which would help them improve their own programs.

The present report contains additional findings on the conditions in which the various curriculum programs are used, and the program teams will no doubt study them.

Other Audiences

As indicated above, the evaluation team strived to satisfy the information needs of various groups interested in mathematics teaching. The Pedagogical Secretariat of the Ministry of Edu-

cation, and the institutes providing in-service education for teachers were additional target audiences, and the representatives of these bodies expressed their interest in meeting the evaluation team and discussing with it the implications of the results for further action.

One of the important target audiences of the study was also the general public.

The evaluation team intends to release a communique for the mass-media only after those interested parties in charge of various aspects of mathematics teaching in the country are given the opportunity to react to the findings. It is not expected that the general public will directly utilize the results of the study, but the reactions conveyed through the mass-media will certainly affect decisions related to mathematics teaching in the elementary school.

Finally, the reaction of the community of educational researchers to the study report will be observed for two reasons. First, their reaction may have consequences for determining the parameters of similar studies in the future. Secondly, attention given to the results of a study by the community of researchers, and their public comments on the validity and significance of the findings, have a great impact on those who are responsible for direct action, and thus reduces the danger that decision makers will disregard the recommendations of the research report.

Increasing the Utilization of the Results

In spite of activities aimed to ensure the utilization of the results, we observed a considerable discrepancy between the expected and the actual level of utilization. One should ask whether this discrepancy can be reduced and what are the most efficient modes for reducing it.

Patton (1985) stated that,

> *... evaluators committed to enhance utilization have a responsibility to train decision makers and information users in evaluation processes and the uses of information. By training stakeholders in evaluation methods and processes, the evaluator is looking to both short-*

> *term and long term utilization. Making decision makers more sophisticated about evaluation can contribute to greater use of evaluation data and evaluation processes over time* (p. 1749).

Patton added that carrying out such training involves financial resources and staff time which are far from being trivial factors. Any attempt to increase utilization should not disregard these recommendations.

In examining the scope of utilization one should pay attention to its definition. Employing a narrow sense of the concept may lead to an underestimation of the actual level of utilization. In his early definition, Alkin (1970) conceptualized utilization as the making of decisions in order to select one of the possible ways of action. In contemporary professional literature utilization is defined in broader terms.

Cousins and Leithwood (1986) introduced a broader definition of the concept of utilization. According to their view, the education of decision makers should also be included in the definition of utilization. They also noted:

> *More recently an even more basic conception of evaluation use has been described: that the mere psychological processing of evaluation results constitutes use, without necessarily informing decisions, dictating actions, or changing thinking* (p. 332).

We felt, however, that the definition suggested by Cousins and Leithwood (1986) is vague and can hardly be used in empirical studies. It might be extremely difficult to collect empirical data for substantiating hypotheses which deal with such a broadly defined concept. Therefore, we prefered to adopt the definition suggested by Patton (1985). He also viewed utilization as a variety of activities, but restricted the domain of relevant activities to those which have observable consequences for the target of the evaluation. Examples of such consequences are: making minor adjustments in programs, reducing uncertainty, enlarging the repertoire of option, increasing control about program activities, and so on.

Utilization of evaluation results may be increased by, first, compiling a list of utilization alternatives, determining what are the conditions which may facilitate utilizations of various types,

and then investing efforts to create such conditions.

This implies that the work of the evaluation team goes beyond submitting a report. Quite frequently it is only an interim stop on the long road of promoting utilization.

Chapter Eight
Summary and Recommendations

Summary

The study described in this report represents an attempt to provide a comprehensive picture about the state of the art of mathematics teaching-learning in the fourth-grade of Israeli Elementary schools. It constitutes one instance in a series of studies designed to provide information on the state of the art of teaching-learning in various subject areas in the Israeli schools. The Israel Curriculum Center together with the Pedagogical Secretariat of the Ministry of Education and Culture worked out the outlines of a multiphase evaluation project, which in a cycle of five-six years, is supposed to provide a comprehensive picture of conditions, methods, and outcomes of teaching-learning in major subject areas. The project was launched in response to an increased demand for accountability in individual schools, subsystems of schools (like those of a single municipality, or a single administratve district of the Ministry of Education and Culture, etc.), and in the whole national system of schools. While the Israel Curriculum Center gave priority to information relevant to the availability of adequate instructional materials, and to the relative effectiveness of various sets of instructional materials, the Pedagogical Secretariat was more interested in obtaining information about the operation of the program delivery system in the country and the actual level of scholastic achievement in mathematics of fourth-grade elementary school pupils in Israel.

Conflicting results reported by various researchers conducting surveys on partial samples of the target population had increased the uncertainty about the achievement levels of pupils and had motivated the Ministry of Education and Culture to conduct national surveys based on a full and representative random sample of the whole target population.

The study described here followed some patterns established by the National Assessment of Educational Progress in the United States (Tyler, 1985) and by the Assessment of Performance Unit in the United Kingdom (Pring, 1981). But the major features of the study were determined on the basis of an analytical review and criticism of the studies of the International Association for the Evaluation of Educational Achievement (IEA) based at the University of Stockholm, Sweden (Postlethwaite, 1985). Israel had previously participated in several cross-national comparative studies of the IEA (see Lewy et al, 1978), and though the present study did not constitute a part of an international study and did not deal with cross-national comparisons, its overall framework reflects a modified version of the IEA study designs.

The study was outlined when the Ministry of Education made operational steps for establishing a Feedback System with the aim of systematically and continuously providing information input for decision makers. At the completion of this research summary the desired parameters of the Feedback System had not yet been determined, and a special committee appointed by the Director General of the Ministry of Education is still in the process of preparing relevant recommendations. The writer of this report relied on the recommendations of a previous committee appointed by the Director General of the Ministry of Education and Culture in 1978. That committee issued a report which recommended that within the Feedback System priority should be given to achievement data, and that achievement studies should be carried out on the basis of data collected from national samples of the whole student population (The Henrietta Szold Institute, 1981). The present study followed these two recommendations. In completing this summary the writer strove to provide an illustration of one possible report-form, which may be prepared within the framework of a national Feedback System. Increased interest in national Feedback System may render this report a useful illustration not only for the Israeli School System but also for other countries. Indeed, the writer of this report, being a faculty member of one of Brazil's universities, intends to present the whole study she did as an illustrative example for the educational system of Brazil.

A Multipurpose Study Responding to Multiple Clientele

The model suggested can be applied to a multipurpose study which is expected to answer a variety of questions, among them those which directly relate to issues of accountability as well as the

relative merits of alternative programs. Moreover, the Feedback System will store information for questions which may emerge in the future. It would serve as an information base for studying trends and changes in the conditions and outcomes of teaching-learning, and it would also facilitate the study of the degree of stability in the relationship between conditions and outcomes of teaching.

The data-base stored in the Feedback System is expected to serve a multiple clientele (frequently and incorrectly referred to as audience). At the whole system level it will provide information for the Pedagogical Secretariat, for the body of supervisors, for Pre-Service and In-Service Teacher Training Institutions and Agencies, for curriculum development teams, for the Educational Commission of the Parliament, and for the general public. At a local level it will provide information for local and district authorities, for individual schools and even for teachers about the status of their classes and of individuals in their classes.

The Feedback System is also supposed to assume responsibility for the dissemination of results available and for increasing the utilization of the data and results across the country.

The present report contains information about the intended curriculum, the implemented curriculum and the attained curriculum in grade four mathematics.

The Intended Curriculum

The intended curriculum is embodied in the National Syllabus and in the instructional programs (textbooks and instructional materials of other types) used in the system. In 1972, a new National Syllabus for elementary school mathematics was produced, which replaced the former syllabus of 1956. In contrast to the old syllabus, which focussed on arithmetical computation, the new syllabus reflects the ideas embodied in the innovative mathematics curricula developed mainly in the U.S. and the U.K. during the 1950's and 1960's. It emphasizes: (1) logical and formal thinking; (2) rules, concepts and relationships; (3) the comprehension of the structure of mathematics as a field of disciplined inquiry; and (4) recommends the utilization of manipulative learning aids.

The study revealed that a high quantity of alternative instructional programs are used in the country. The principals in a

sample of 272 schools reported the utilization of 22 different sets of instructional materials. Since some of these sets were mentioned only in one school, one may assume that in the total system more than 22 different programs are used. This finding surprised experts dealing with mathematics teaching in the country. It was even more surprising to find out that some of these programs were entirely unfamiliar to the Chief Superintendent for Mathematics Teaching, and that some of the programs appear on a list of those instructional materials rejected by the Ministry of Education and Culture, and whose use is not allowed.

Three of the instructional programs listed by the principals are used in 68 percent of all classes in the country. These books represent the innovative teaching of mathematics. Two of them were prepared within the framework of institutions (The "Simple Math" program by the Instructional Television in Israel, and the "One, Two and Three..." program by the Institute of Educational Technology along with the Israel Curriculum Center), and the third one ("New Mathematics") was designed by a distinguished primary school teacher (Mr. Moshe Hershko), who pioneered the introduction of the ideas and teaching of the 1960's new mathematics into Israeli schools. The three most widespread programs used in Israel differ from each other with regard to the following issues: utilization of manipulative teaching aids, reliance on the child's immediate environment, the proportion of word problems and their place in the sequence of contents taught. But beyond these differences all of three these programs reflect the innovative approach specified in the National Syllabus.

The Implemented Curriculum

In dealing with the implemented curriculum we focussed on: (1) the conditions of implementations; (2) the teachers' actual involvement in activities related to implementing the program; and (3) the coverage of the curriculum content.

(1) Conditions of implementation

We examined various conditions related to studying mathematics in elementary schools. First, we examined home conditions which are related to mathematics learning. It was found that parents attribute great importance to studying mathemathics and to obtaining high grades. The students internalize these attitudes toward mathematics and they perceive mathematics as one of the

most important subjects in school. Moreover, they list mathematics among the most liked subjects. It is difficult to interpret the meaning of such statements. In light of the widespread concern about mathematics-phobia or mathematics-anxiety of school children, one may wonder whether these statements represent the sincere feelings of pupils toward mathematics, or alternatively, they submit themselves to a socially desirable posture and pretend to enjoy the subject. It should be noted that in several studies carried out in Israel we also found students indicating that they like mathematics. It seems advantageous to study this issue more thoroughly, and to collect less unobtrusive data and less reactive to the conditions of eliciting responses.

Other aspects touched on time dedicated to mathematics teaching in grade four. The national school time-schedule allocates four weekly hours to mathematics study in the school. In reality, however, 75 percent of the schools allocate five weekly periods. This is further evidence of the significance attributed to mathematics teaching-learning within the framework of schools.

The third aspect of the conditions relates to the teachers' professional education in teaching mathematics. The data revealed that 57 percent of the teachers received professional training in the area of teaching mathematics in Teacher Training Institutes, and 73 percent of the teachers participated in in-service training activities related to the particular program which he/she uses for teaching in class.

This is a relatively high figure, since elementary school teachers usually receive a general professional education, which touches only marginally on problems of teaching mathematics. It seems that mathematics, and mainly innovative programs in mathematics, are more likely to engender mathematics-phobia among teachers than among pupils.

(2) The Teacher's Actual Involvement

The second aspect of implementation examined in the study dealt with the teachers' actual involvement in activities which are components of implementation. It was found that 90 percent of the teachers claim to be familiar with the National Syllabus, though only 60 percent of the teachers use the syllabus constantly. About 50 percent of the teachers use the Teacher's Guides. This is a relatively low figure, and it is explained by the fact that for several instructional programs no Teacher's Guide has yet been

produced. This is the case with the wideley-used fourth-grade mathematics program of the "Simple Math". Another useful piece of information derived from the data is the fact that 63 percent of the teachers participated in school conferences dealing with the problems of teaching mathematics (44 percent of the teachers did so more than one time during the year).

The professional guidance services provide support for approximately 50 percent of the teachers.

A noteworthy finding of the study is that 60 percent of the teachers employ an expository teaching style (labelled in Israel as frontal teaching), where a single teacher explains things to all pupils in the class, and all pupils work at a given period of time on the same assignment. It can be said that in most components of the implementation process, as well as under the implementation conditions, we find that about 50 percent of the schools achieve a positive score and 50 percent a negative one. Thus, for example, 50 percent of the teachers received a professional education for teaching mathematics, 50 percent of them use a Teacher's Guide, and 50 percent benefit from guidance services. With regard to other several implementation components we find better results. Thus, for example, 63 percent of the teachers participated in school conferences which deal with matters of teaching mathematics. But this figure is balanced by a low performance figure with respect to using advanced class organization patterns: sixty percent of the teachers still adhere to the conventional expository teaching mode.

This does not mean that in only 50 percent of the schools did the new mathematics programs have an impact. It is rather that one may observe a great variation between schools regarding the level of implementation.

Approximately 15 percent of the schools are high implementers, adopting a relatively high proportion (70-80 percent) of implementation components. About 65 percent of the schools are moderately high implementers, and only 10 percent of the schools are relatively low implementers of the innovative mathematics teaching approach.

(3) <u>The coverage of the curriculum content</u>

A major implementation measure in this study was the Opportuni-

ty to Learn (OTL) variable. Teachers were asked to mark whether the pupils in their classes had an opportunity to learn curriculum contents listed in the syllabus and which are represented in the test items. They were asked to mark their responses to two lists. One list contained 21 topics which are recorded in the National Syllabus, and the other contained 40 items of the achivement test used in the study. The mean percentage on the list of topics was 53.3 percent (standard deviation 13.5), and on the list of the test items 87.2 percent (standard deviation 11.5). Thirty percent of the teachers indicated that the pupils had an opportunity to learn all curriculum contents which were represented in the test items, and only 5 percent of the schools indicated that they covered all topics appearing in the National Syllabus.

There is a considerable difference between the classroom coverage of the test items (87.2 percent) and the classroom coverage of the National Syllabus topics. A possible reason for this discrepancy is that the developers of the test sampled only those topics which, as far as they knew, are taught in most schools. They did not construct test items which examined knowledge of topics usually taught only during the last quarter of the school year. In other words, it can be said that the test developers focussed on those curricular contents which, according to their familiarity with the state of the art of teaching mathematics in the schools, are actually covered. Using more formal terms, the National Syllabus constitutes the intended curriculum as defined at the system level, and the test items represent a compromise between the nationally prescribed curriculum and the "desired curriculum" as defined by mathematics educators. These educators maintain continuous contact with teachers and schools, and drawing on their observations and experiences, they were satisfied with a more moderate scope of accomplishment than that specified in the national curriculum. The "desired curriculum" constitutes a reduced set of the intended curriculum as prescribed in the National Syllabus. It was implemented in the system to an extent of 87 percent, while the National Syllabus was implemented only to 53 percent.

The Attained Curriculum

The outcomes of studying mathematics has been defined as the attained curriculum. It was measured by a 40 items multiple-choice type achievement test constructed specially for the present study. It contained a substantial proportion of item (21 items) relating to contents covered in grade three, which examined the mastery of

skills and contents taught in previous years. The items contained computational skills and also word problems. Ten items of the test were defined as items measuring Minimum Competency Requirements. Two standards of mastery on the Minimum Competency test were defined: a severe standard according to which scores of 9 and 10 were considered as pass-scores, and a more lenient standard where scores of 7 and 8 were considered as passing. The data analysis was carried out both at the individual level and at the class level (since each schools was represented by a single class, in this case there is an identity between the school level and class level analysis). It was decided to carry out a class level analysis, since the class is an ecological unit and pupils studying in a single class are exposed to the same within-class teaching-learning activities. An additional reason for carrying out class level analysis was the fact that the Israeli elementary school usually has a "catchment area" of a single neighborhood, and therefore the socio-economic differences across schools represent socio-economic differences across residential areas within the cities, and between cities and townships. Indeed, the etha coefficient characterizing the homogeneity of schools (or the between-school achievement differences) is .26, which is a relatively high index, and provides further justification for carrying out class or school level analysis.

The major findings related to the attained curriculum are summarized as follows:

1. The achievement level on the grade three curricular materials was considerably higher than on grade four curricular materials (means are 83 and 65 respectively).

2. The variation across individuals and across schools is smaller with regard to achievements on the grade three curricular units than on the grade four curricular units (individual score's standard deviation is 15.8 vs. 21.4, and the class level standard devitions are 7.1 vs 12.4 respectively).

3. No substantial differences were found between achievements on word problems and computational skills.

4. The achievement in geometry and knowledge of fractions was considerably higher than on other areas of mathematics knowledge tested.

5. On the Minimum Competency test, 77 percent of the pupils passed the severe standards of the test (i.e. 9 or 10 correct responses on a 10 item test), and approximately 95 percent of the pupils had a score of 7 or more on that test.

6. Using the intermediate standard (more than 7 correct responses) for the Minimum Competency test, it was found that in 30 percent of the schools, none of the pupils failed this test, and only in approximately 15 percent of the schools were there more than three pupils who failed the test. One may say that these schools are in need of high compensatory education support.

The Utilization Aspect

Communication about the results of the study have been made through three different channels: personal communications, conferences and interim reports. Through these three channels, the communication was directed to various clienteles or stakeholders of mathematics teaching-learning: curriculum development teams; the bodies of superintendents and supervisors; school principals and teachers. We will report here briefly about the nature of the communication to the above-mentioned groups.

Curriculum development teams - Personal communication was established with the curriculum development teams at the phase of planning the study and developing the instruments. At the completion of the first comprehensive report of the study, a conference was conducted at the Israel Curriculum Center, which was attended by the mathematics curriculum teams and by members of other curriculum teams working at the Center. The discussions focussed on problems of evaluating curriculum implementation and outcomes in general, and the payoff of the discussion touched on methodological issues of curriculum evaluation, as well as on problems observed in relationship to the fourth-grade mathematics programs. The teams were interested in finding out whether some programs function better than others. Their approach was rather competitive and wanted to know what status their program had in comparision to other programs. It was quite difficult to eliminate these competitive comparisons.

The superintendent and the supervisors - The superintendents

and the supervisors were interested in the way teachers handle various sets of instructional materials (instructional programs). Quite unexpectedly they were also interested in the pupils' achievement level in each particular school. The comparison of achievements was a dominant focus of their interest, but they wanted to compare achievements in various schools rather than programs. The personal contact with the Chief Superintendent of mathematics teaching was continuous and interim results were communicated to him as soon as they became available. A conference with the Tel-Aviv District School Supervisors was conducted after the publication of the first comprehensive report.

The school principals and teachers - School principals and teachers were briefed a short time after the test and the pupil's questionnaire were conducted in their schools. An effort was made to provide them information at a time when they were able to use it for the purpose of grading the students. Thus, results obtained in a particular school were released before the grading conferences took place.

A written document was prepared for the teachers which contained information about achievement at the system level, and information related to achievement in their own school. Principals and teachers were informed about achievement in other schools only in general terms, and these schools were not identified.

A national conference of teachers and principals was conducted approximately three months after the administration of the test, at a time when the results were already available.

It is of interest to note that teachers revealed an interest in validating their grades and looked for means to create homogeneous ability groups. They payed little attention to the diagnostic value of the findings for improving their teaching or to adapt their teaching to the particular needs of their students.

Conditions promoting utilization - The activities described above represent attempts to communicate the results to the stakeholders of the evaluation object. At this phase, little can be said about utilization and one has to be satisfied with making comments on the climate of the communication and on its effectiveness. It can be said that the climate was very favorable and the stakeholders revealed great interest in the findings of the evaluation. But this does not necessarily mean that action will be taken on issues which have been highlighted by the e-

valuation findings. One should remember that in a complex organization, such as an educational one, there are conflicting interests, and powerful administrators may feel quite comfortable in disregarding evaluation recommendations. After all, an organization does not consist of a group of altruistic managers, who are willing to sacrifice their power and authority when evaluation results suggest that curtailing the scope of their action may be beneficial to attaining organizational goals. One cannot easily specify the range of actions which may be useful for promoting utilization of evaluation results, but it may be definitely stated that producing an evaluation report in iteself does not ensure the utilization of the results. The evaluators have to be devoted in motivating decision makers to consider evaluation findings. And whenever evaluators do this with ingenuity, they are likely to succeed.

While recommendations for action were communicated to decision makers through other channels, it seems appropriate to list in the concluding section of this study some recommendations directed to various stakeholders of mathematics instruction in Israel, and to professional evaluators who carry out evaluation and assessment studies in educational systems of any kind.

Recommendations

Recommendations to Producers of Instructional Programs

1. Despite the systematically planned and conducted formative evaluation of instructional materials by the Israel Curriculum Center, a substantial proportion of teachers still complain that they find it difficult to comprehend instructions contained in the textbooks. While there are differences in the scope of such complains between the textbooks examined in this study, it cannot be overemphasized that the comprehensibility of texts shoud be meticuluously evaluated on representative samples of the intended users. Special signs used to mark sections which contain more complicated texts may be a useful device for reducing teachers' frustration.

2. Systematic review exercises of contents and skills taught in lower classes should be spread over the textbooks. The producers of instructional materials should not delegate to

teachers the task of finding review exercises, but rather they should facilitate the teachers' work by providing opportunities for review material in the textbooks.

3. Most teachers, despite their use of the Teacher's Guide, are interested in planning the teaching of broad topics and sections of the program, rather than in the detailed planning of the lessons day after day. This fact should be considered in writing the Teacher's Guides. Authors of the Guides should pay great attention to introducing broad topics rather than to minute details related to a particular lesson.

Recommendations for Planners of Teacher Education Programs

4. It seems almost trivial to suggest that teacher education should pay attention to special problems of teaching elementary school mathematics. Nevertheless, it should be stated that the general component of the teacher education program should have a strong mathematics-specific component. The study revealed that about 50 percent of fourth-grade mathematics teachers did not have any specialized training for teaching mathematics. It is likely that this situation will not change in the near future, and most experts believe that it is undesirable to delegate the task of mathematics teaching in the fourth grade to specialized teachers. Therefore, increasing the mathematics-specific elements in the general training program seems to be the most practical way for improving the teachers' professional knowledge in the field of mathematics teaching.

Recommendations to the National and Regional Educational Authorities

5. Approximately 5 percent of the pupils failed to pass even the most lenient standards on the Minimum Competency test, and 11 percent of the pupils did not pass the intermediate level standard on that test. In approximately 15 percent of the classes across the country there is a high concentration of pupils who failed the Minimum Competency test (more than 3 pupils in the class).

The national and regional authorities should identify those schools which have a relatively high concentration of

pupils failing the Minimum Competency test and provide a compensatory education budget for dealing with the "weak" students within the framework of their regular budget.

6. At grade four of the elementary school no differences were observed between the achievement of boys and girls. Nevertheless, at higher grade levels in the secondary school there is a considerable difference between the percentage of boys and girls studying mathematics (the percentage of boys is higher than that of girls).

   School authorities should identify the girls who excell in mathematics and try to encourage them to study extended course options in mathematics.

Recommendations to Superintendents and Supervisors

7. There is a need to examine why some schools utilize instructional programs and/or isolated textbooks which do not appear in the list of the officially approved textbooks. This does not imply that central authorities should forcefully eliminate the utilization of such programs and/or books from classes. It may well be that the decisions of the instructional programs and/or textbook-approving authorities need to be modified, or that the whole procedure of approving textbooks should be modified or even fully abolished.

8. Superintendents and supervisors should facilitate the dissemination of the National Syllabus in schools, and should motivate teachers to familiarize themselves with the syllabus and use it for the purpose of instructional planning.

Recommendations to Principals and Teachers

9. School principals and teachers have to select instructional programs or textbooks in various subject fields as well as in mathematics. According to the results of the present study, not one of the three leading instructional programs was found to produce better achievements than the others. The comparison of the achievements of users of these programs suggests that they function equally well, and the principals and teachers should base their decisions on personal preferences and local considerations.

10. Teachers should participate in the decision making about the selection of instructional programs and isolated textbooks and/or other types of teaching materials. The findings of the study revealed that teachers are dissatisfied with the procedures of instructional program selection and want to have more autonomy in these matters. It may well be that this is not merely a within school issue and there is a need for system-wide changes. Surprisingly, the school regulations do not contain any instructions concerning the selection of instructional programs (textbooks and specific instructional aids). For a very long time no substantial differences existed among various sets of textbooks and instructional aids, and in several important subject areas no alternative programs of teaching-learning existed, and therefore the selection of sets of instructional materials was not considered a crucial issue.

    In the last two decades we experienced a great increase in the number of isolated textbooks and particular teaching aids published in Israel. The present survey revealed that 22 different instructional programs are used in grade four. These programs represent different epistemological and psychological views about the nature of matehmatics and teaching-learning the subject, and consequently, more attention should be given to the selection of any type of instructional materials. Teachers need to obtain more information about the viewes and approaches being represented in the specific textbooks and instructional aids.

11. A substantial proportion of schools do not conduct conferences and teachers' meetings, which are dedicated to problems of teaching mathematics. Schools should pay more attention to these problems, and it is highly recommended that this subject receive systematic treatment.

12. Teachers and principals often encounter difficulties in using evaluation results for diagnostic purposes. It has been shown that test results have implications for the adapting of school curriculums to local needs, and that they help provide individually corrective teaching for those who need it. Schools should develop strategies which may facilitate the diagnostic use of the test results.

13. Most pupils indicate that they like mathematics. This finding contradicts the widespread belief that a large proportion of

children suffer from mathematics-phobia. It seems worthwhile to examine to what extent is mathematics-phobia widespread among school children in Israel.

Due to discrepancy between the findings of surveys and widespread beliefs about the attitude of school children toward mathematics, it may be advisable to collect information about attitudes toward mathematics not only through questionnaires, but also through less obtrusive measures, such as depth interview, observation of behavior, etc.

Recommendations to Evaluation Professionals

14. The successful dissemination of the evaluation findings is a precondition of its effective use in various contexts. The evaluator has to make arrangements so that the evaluation findings should be communicated to all stakeholders of the issue examined. It is not enough to mail the report to prespecified addresses. Successful dissemination may require vigorous action on behalf of the evaluator, who may have to maintain personal contact with decision makers and use suitable strategies of other types to elicit the reaction of the stakeholders to the findings.

Since the work of the evaluator is not over with the production of the report, it is necessary to make contractual arrangements about the post-publication work to be carried out. Experience suggests that this can be better done if those who commission the evaluation study have on-going contact with the evaluator, either in the form of a contract for a series of evaluaton studies or in the form of institutional dependence (i.e. both the production and the evaluation unit are administratively and financially subordinated to the same higher level of authority).

15. Evaluators should be cautioned to limit the scope of their data collection to the amount of data which they are able to process and to summarize. The present study was based on a relatively small quantity of data. The pupils were given a few attitude and biography related questions and a 40-item achievement test. Despite the extensive data analysis and summary in this report, one cannot say that all information which can be derived from the data has been surveyed in this report. Indeed, the evaluator had to make decisions on what

kind of data analysis should be carried out, and on the basis of obtained-data summaries she had to make decisions what should be communicated in the reports. Secondary analysis of data may lead to the production of additional, useful information.

The report presented here offers evidence that a single set of data may be effectively used for producing information which has bearing both on assessing the achievement of individuals, and at the same time on evaluating the relative worth of various programs and ways of action.

Bibliography

Adar,L. et al Teaching in heterogeneous classes: Report of the bi-national Israel-USA seminar. Jerusalem: 1981 (mimeogr.)

Aiken,R.L. Mathematics, attitude towards. In: T. Husen and T.N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Airasian,P.W. and Madaus,G.F. A Post-hoc technique for identifying between program differences in achievement. Studies in Educational Evaluation, 1978, 4, 1-8.

Algabrabi,M. Indices for characterizing schools, and a method for allocating funds to schools. "Megamoth", 1975, 21, 219-227.

Alkin,M.C. Evaluation theory development. Evaluation Comment, 1969, 2(1), 2-7.

Alkin,M.C. and Daillak,R. Evaluation studies, impact of. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Alkin,M.C., Daillak,R. and White,P. Using evaluation: does evaluation make a difference?. Beverly Hills, Calif.: Sage, 1979.

Anderson,L.W. Opportunity to Learn. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Atkin,M. Case studies in science education. In: P. Tamir (ed.) Curriculum Implementation. Jerusalem: Hebrew University, Science Teaching Center, 1979.

Baker,E. and Choppin,B. Minimum competency testing. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Bar,R. Content coverage in classrooms. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Bashi,J., Kahan,S. and Davis,D. Educational achievement in the Arabic Elementary School in Israel. Jerusalem: Hebrew University, School of Education, 1981 (in Hebrew).

Ben-Peretz, M. Curriculum analysis as a tool of evaluation. In: A. Lewy and D. Nevo (eds.) Evaluation Roles in Education. New York: Gordon and Breach, 1981.

Berk,L. Back to basics movement. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Berk,R.A. A consumer's guide to setting performance standards on criterion-referenced tests. Review of Educational Research, 1986, 56, 137-172.

Berliner,D.C. Tempus educare. In: P.L. Peterson and H.J. Walberg (eds) Research on teaching: Concepts, findings and implications. San Francisco: McCutchan, 1979.

Block,J.H. Mastery Learning: theory and practice. New York: Holt, Rinehart and Winston, 1971.

Block,J.H. Standards and criteria: a response. Journal of Educational Measurement, 1978, 15(4), 291-295.

Bloom,B.S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Bloom,B.S. Tryout and revision of educational material and methods. In: A. Lewy (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.

Bloom,B.S., Engelhart,M.B., Furst,E.J., Hill,W.H., and Krathwohl,D.R. Taxonomy of educational objectives, handbook 1: cognitive domain. New York: McKay, 1956.

Bloom,B.S., Hastings,J.T. and Madaus,G.F. Handbook on formative and summative evaluation of student learning. New Jersey: McGraw-Hill, 1971.

Boruch,R. On evaluation policy in Israel and USA. In: A. Lewy and S. Kugellmass (eds.) Decision oriented evaluation in education: the case of Israel. Philadelphia: International Science Services, 1981.

Boruch,R. Experimental tests in education: recommendations from the Holtzman Report. The American Statistician, 1982, 36, 1-14.

Brodinsky,B. Back to basics: the movement and its meaning. Phi Delta Kappan, 1977, 58, 522-27.

Browder,L.H. Accountability in Education. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Bruner,J.F. The process of education. New York: Vintage, 1960.

Burton,N.W. Societal standards. Journal of Educational Measurement, 1978, 15(4), 263-271.

Butcher,J.E. Comparison of effects of distributed and massed problem assignments on the homework of grade nine algebra students. Ph.D. Dissertation, Abstract Intern. 76-8683. New Jersey: Rutgers University, 1975.

Campbell,D.T. Reforms as experiments. American Psychologist, 1969, 24(4), 409-429.

Campbell,D.T. Experiments as arguments. In: E.R. House et al (eds.) Evaluation studies review annual (Vol. 7). Beverly Hills, Calif.: Sage, 1982, 117-127.

Campbell,D.T. and Fisher,D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Carlson, D. Duplex designs: test designs serving the purposes of both individual measurement and curricular assessment. Paper presented at the International Association for Educational Assessment Conference in Blantyre (Malawi), June 13-17, 1983.

Carroll,J.B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Carroll,J.B. The teaching of French as a foreign language in eight countries. New York: Wiley, 1975.

Carroll,J.B. The IEA studies in English and French as foreign languages: Implications for educational planning. Studies in Educational Evaluation, 1977, 3(1), 1-5.

Charters,W.W. and Jones,J. On the risk of appraising non events in program evaluation. Educational Researcher, 1973, 2(11), 5-7.

Chen,M., Lewy,A. and Erlich,O. The relationship between class size with pupils' achievement in fourth grade. Magamoth, 1971, 18(2), 247-9.

Chew Tow Yow. Evaluation at the planning stage. In: A. Lewy (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.

Choppin,B. Evaluation, Assessment and Measurement. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Cohen,D.K. Loss as a theme in social policy. Harvard Educational Review, 1976, 46, 553-71.

Cohen,D.K. Curriculum decision making, school based. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Cohen,D.K. and Haney,W. Minimums, competency testing and social policy. In: R.M. Jaeger and C.K. Tittle (eds.) Minimum Competency Testing. Berkeley, Calif.: McCutchan, 1980.

Cohen,J. and Cohen,P. Applied multiple regression/correlation analysis for the behavioral sciences. New Jersey: John Willey, 1975

College Entrance Examination Board. On further examination: report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. New York: College Entrance Examination Board, 1977.

College Entrance Examination Board. National college bound seniors 1980. New York: CEEB, 1980.

Coulter,F. Homework: A neglected research area. British Educational Research Journal, 1979, 5, 21-33.

Cousins,J.B. and Leithwood,K.A. Current empirical research on evaluation utilization. Review of Educational Research, 1986, 56, 311-364.

Cronbach,L.J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.

Cronbach,L.J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 26, 867-876.

Cronbach,L.J. Designing evaluation of educational and social programs. (2nd printing) San Francisco, Calif.: Jossey-Bass, 1983.

Cronbach,L.J., Ambron,S.R., Dornbush,S.M., Hess,R.D., Hornik,R.C., Phillips,D.C., Walker,D.E., and Weiner,S.S. Toward reform of program evaluation. San Francisco, Calif.: Jossey-Bass, 1980.

Cunningham,W.G. The National Assessment: a critical review and a new direction. The Clearing House, 1976, 50, 82-7.

Dar,Y. Sex differences in school achievements of adolescents borned in "kibutzim". Tel-Aviv, The Institute of Social Research of the "Kibutzim" Confederation, 1974.

Dar,Y. and Resh,N. Classroom composition and pupil achievement: A study of the effect of ability-based classes. New York: Gordon and Breach, 1986.

Davis,D. Evaluating the family as an educational framework. In: A. Lewy and D. Nevo (eds.) Evaluation Roles in Education. New York: Gordon and Breach, 1981.

Davis,E. Teachers as curriculum developers. Sydney: Allen and Unwin, 1981.

Davis,R.B. Mathematics: elementary-school programs. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Denzin,N. The research act. New York: McGraw-Hill, 1978.

Dunn,W.N. Reforms as arguments. In: House,E.R., et al (eds.) Evaluation studies review annual (Vol. 7). Beverly Hills, Calif.: Sage, 1982, 83-116.

Eden,S. Implementation of innovations in education. Studies in Educational Evaluation, 1979, Monograph no. 2.

Education Commission of the States. The third national mathematics assessment: results, trends and issues. Denver: National Assessment of Educational Progress, 1983.

Eisner,E.W. On the uses of educational connoisseurship and criticism for evaluating classroom life. Teachers College Record, February 1977, 78(3), 347-360.

Eisner,E.W. The educational imagination: on the design and evaluation of school programs. New York: Macmillan, 1979.

Eraut,M.R. In-service education for innovation. London: National Council for Educational Technology, 1972.

Farr,R. and Olshavsky,J.E. Is minimum competency testing the appropriate solution to the SAT decline? Phi Delta Kappan, 1980, 61, 528-30.

Fey,J.T. Investigating the minimal competency trend around the world (1976-1980). In: R.J.Riehs (ed.) An international review of minimal competency programs in mathematics. International Study Group on Minimal Competencies in Mathemat- ics, Second International Mathematics Study - International

Association for the Evaluation of Education Achievement (IEA). Columbus, Ohio: The Eric Science, Mathematics and Environmental Clearinghouse, Center for Science and Mathematics Education (The Ohio State University): March 1981.

Finley,C.J. and Berdie,F.S. The National Assessment approach to exercise development. Ann Arbor, MI: The National Assessment Staff Offices, 1970.

Fisher,C.W. et al Teaching behaviors, academic learning time, and student achievement. In: C. Denham and A. Liberman (eds). Time to learn. Washington, DC: National Institute of Education, 1980.

Fleming,S.M. The effect of class size on pupils' achievements. "Megamoth", 1959, 10(4), 1959.

Forbes,R.H. National assessment: a source of descriptive educational data. In: G.R. Austin and H. Garber (eds.) The raise and fall of national test scores. New York: Academic Press, 1982.

Fraser,B.J. Directions in curriculum evaluation. Studies in Educational Evaluation, 1984, 10, 125-134.

Fullan,M. The relationship between evaluation and implementation. In: A. Lewy and D. Nevo (eds.) Evaluation roles in education. New York: Gordon and Breach, 1981.

Fullan,M. The meaning of educational change. Canada: The Ontario Institute for Studies in Education, 1982.

Fullan,M. and Pomfret,A., Research on curriculum and instruction implementation. Review of Educational Research, 1977, 47(1), 335-97.

Gagne,R.M., Mayor,J.R., Garstens,H.L. and Paradise,N.E. Factors in acquiring knowledge in mathematical tasks. Psychological Monograph, 1962, 76(526).

Gage,N.L. The scientific basis of the art of teaching. New York: Teachers College, 1978.

Gall,M.D. Handbook for evaluating and selecting curriculum mate-

rials. Boston: Allyn and Bacon, 1981.

Galor,N. Study of low achievers in history by junior high school students. Unpublished Master's thesis. Tel-Aviv: Tel-Aviv University, School of Education, 1983.

Getzels,J.W. and Jackson,P.W. The teacher's personality and characteristics. In N.L. Gage (ed) Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Glaser,R. Instructional technology and measurement of learning outcomes. American Psychologist, 1963, 18, 519-21.

Glass,G.V. Standards and criteria. Journal of Educational Measurement, 1978, 15, 237-261.

Goodlad,J.I. and Klein, M.F. Behind the classroom door. Washington Ohio: Jones, 1970.

Gorth,W.P. A study of minimum competency programs. Amharst, MA: National Evaluation Systems, 1979 (ERIC doc. ED 185 123).

Gorth,W.P. and Perkins,M.R. A study of minimum competency testing programs: final typology report. Amharst, MA: National Evaluation Systems, 1979 (ERIC doc. ED 185 125).

Greenbaum,W., Garet,M.S. and Solomon,E.R. Measuring educational progress: a study of the National Assessment of Educational Progress. New York: McGraw-Hill, 1977.

Guba,E.G. The failure in utilizing the results of evaluation. In: C.H. Weiss (ed.) Evaluating action programs: readings in social action and education. Boston: Allyn and Bacon, 1972.

Guba,E.G. Toward a methodology of naturalistic inquiry in educational evaluation. CSE Monograph Series, No. 8, 1978.

Guba,E.G. and Lincoln,Y.S. Effective evaluation. San Francisco: Jossey-Bass, 1981

Hambleton,R.K. On the use of cut-off scores with criterion-referenced tests in instructional settings. Journal of Educational Measurement, 1978, 15(4), 277-289.

Hameyer,U. Curriculum Theory. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Haney,W and Madaus,G. Making sense of the competency testing movement. Harvard Educational Review, 1978, 48, 462-484.

Harlen,W. The role of evaluation in teaching. In: A. Lewy and D. Nevo (eds.) Evaluation roles in education. New York: Gordon and Breach, 1981.

Hershko,M. New Mathematics for grade one: textbook Tel-Aviv: Dvir, Co.,Ltda. 1975 (a).

Hershko,M. New Mathematics for grade one: Teacher's Guide. Tel-Aviv: Dvir, Co.,Ltda. 1975 (b).

Hershko,M. New Mathematics for grade two: textbook. Tel-Aviv: Dvir, Co.,Ltda. 1975 (c).

Hershko,M. New Mathematics for grade two: Teacher's Guide. Tel-Aviv: Dvir, Co.,Ltda. 1975 (d).

Hershko,M. New Mathematics for grade three: textbook (2a. edition) Tel-Aviv: Dvir, Co.,Ltda. 1976 (a).

Hershko,M. New Mathematics for grade three: Teacher's Guide. Tel-Aviv: Dvir, Co.,Ltda. 1976 (b).

Hershko,M. New Mathematics for grade four: textbook (2a. edition) Tel-Aviv: Dvir, Co.,Ltda. 1977 (a).

Hershko,M. New Mathematics for grade four: Teacher's Guide. Tel-Aviv: Dvir, Co.,Ltda. 1977 (b).

Hilton,T.L. and Patrick,C. Cross-sectional versus longitudinal data: An empirical comparison of mean differences in academic growth. Journal of Educational Measurement, 1970, 7, 15-24.

Hively,W., Maxwell,G., Rabehl,G., Sension,D. and Lundin,S. Criterion-referenced curriculum evaluation: a technical handbook and a case study from the minnemast project. CSE Monograph Series in Evaluation, No. 1, 1973.

House,E.R. Evaluation theory building. Studies in Educational Evaluation, 1979, 5, 149-150.

House,E.R. Evaluating with validity. Beverly Hills, Calif.: Sage, 1980.

House,E.R. Introduction: scientific and humanistic evaluations. In: E.R. House et al (eds.) Evaluation studies review annual (Vol. 7). Beverly Hills, Calif.: Sage, 1982, 15-25.

House,E.R., Mathison,S., Pearsol,J.A., and Preskill,H. (eds.) Evaluation studies review annual (Vol. 7). Beverly Hills, Calif.: Sage, 1982.

Howson,G., Keitel,C. and Kilpatrick,J. Curriculum development in mathematics. Cambridge: Cambridge University Press, 1981.

Husen,T. International study of achievements in mathematics: a comparison of twelve countries. Stockholm, Almquist and Wiksell, 1967.

Husen,T. Policy implications of the IEA findings and some of their repercussions on national debates on educational policy. Studies in Educational Evaluation, 1977, 3, 129-141.

Hutton,L.A. Calculators in mathematics education. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Instructional Television Center. Simple Math (two textbooks). Tel-Aviv: Instructional Television Center, 1981.

International Association for the Evaluation of Educational Achievement. Second Study of Mathematics. Bulletin 4, December 1979.

Jaeger,M.R. and Tittle,C.K. Minimum competency testing. Berkeley, Calif.: McCutchan, 1980.

Jick,T.D. Mixing qualitative and quantitative methods: triangulation in action. Administrative Science Quartely, 1979, 602-11.

Kant,L. and Orr,L. Profiles, educational. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Keeves,J.P. Profile analysis. In: T.Husen and T.N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.(a)

Keeves,J.P. Sex differences in ability and achievement. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.(b)

Kellagham,T., Madaus,G.F. and Rakow,E.A. Within-school variance in achievement: school effects or error?. Studies in Educational Evaluation, 1979, 5, 101-107.

Kerlinger,F.N. Foundations of behavioral research. (2nd edition) Japan: Holt-Saunders, 1981.

Kerlinger,F.N. and Pedhazur,E.J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, 1973.

Klein,S. Evaluating tests in terms of the information they provide. Evaluation Comment, June 1970, 2(2), 1-6.

Kleinberger,A. Refelexions about equality in education. Megamoth, 1964, 13, 257-288.

Kliebard,H.M. Curriculum movements in the USA. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Kline,M.F., Tye,A.K., Wright,E.J. A study of schooling: Curriculum. Phi Delta Kappan, 1979, 244-248.

Larkins,A.G. Minimum competency test: a negative view. In: H.D. Mehlinger and O.L. Davis (eds.) The social studies 80th yearbook, part 2 (National Society for the Study of Education -NSSE-). Chicago: University of Chicago, 1981.

Leide, G. Expert judgments as evaluation data. In: A. Lewy (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.

Lesh,R. Applied mathematical problem solving. Educational Studies, 1981, 12, 235-264.

Lesh,R. Mathematical applications: Primary schools. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Levin,T. Classroom instructional environment. In: A. Lewy and D. Nevo (eds.) Evaluation roles in education. New York: Gordon and Breach, 1981.

Levin,T. Arithmetic: educational programs. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Lewy,A. The role of evaluation. (in Hebrew) In: S. Eden (ed.) On the New Curricula. Tel-Aviv: Maaloth, 1971.

Lewy,A. Opportunity to learn and achievement in three subject matter areas. The Journal of Experimental Education, 1972, 41(1).

Lewy,A. Discrimination among individuals versus discriminaton among groups. Journal of Educational Measurement, 1973, 10, 19-24.(a)

Lewy,A (ed.) Studies in curriculum evaluation. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1973.(b)

Lewy,A. (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.(a)

Lewy,A. The nature of curriculum evaluation. In: A. Lewy (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.(b)

Lewy,A. Setting up an infrastructure for a multipurpose feedback system on educational achievement. In: U. Trivellato and A. Zuliani: Informazione statistica, gestione della scuola e programmazione scolastica. Venezia: Fondazione G. Cini, 1979.

Lewy,A. The scope of educational evaluation: an introduction. In: A. Lewy and D. Nevo (eds.) Evaluation roles in education. New York: Gordon and Breach, 1981.(a)

Lewy,A. Minimal mathematical competencies in Israel. In: R.J. Riehs (ed.) An international review of minimal competency programs in mathematics. Prepared for the International Study Group on Minimal Competencies in Mathematics, in cooperation with the Second International Mathematics Study - International Association for the Evaluation of Education Achievement (IEA). Columbus, Ohio: The Eric Science, Mathemat- ics and Environmental Clearinghouse, Center for Science and Mathematics Education (The Ohio State University): March 1981.

Lewy,A. Educating teachers to comprehend basic components in innovative programs. In: P. Tamir, Hofstein,A. and M. Ben-Peretz (eds). Preservice and inservice training of science teachers. Philadelphia: Balaban International Science Services, 1983.(a)

Lewy,A. Attitudes and attainments in mathematics:a technical report of a survey in grades 8 and 12. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1983 (in Hebrew with English summary).(b)

Lewy,A. Achievements in Tel-Aviv schools. Tel-Aviv: Tel-Aviv University, School of Education, Unit for Evaluation Research and Measurement, 1984 (in Hebrew).(a)

Lewy,A. Minimum requirement program: a potencial device for promoting equality. Paper presented at the First International Conference on Education in the 90s. (Also published in Studies in Educational Evaluation, 1985, 11, 315-320.) Tel-Aviv: Tel-Aviv University, 1985.(b)

Lewy,A. Systemwide achievement survey. Tel-Aviv University, School of Education, 1985 (mimeograph, in Hebrew).

Lewy,A. Twenty years of evaluation activities at the Israel Curriculum Center. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1987.

Lewy,A. and Alkin,M.C. Impact of a major national evaluation study: Israel's Van Leer Report. Los Angeles: University of

California, Center for the Study of Evaluation, 1983.

Lewy,A. and Chalfon,I. Attitudes toward various school subjects. In A. Lewy (ed.) Studes in curriculum evaluation. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1973.

Lewy,A. and Chen,M. Differences in achievements: a comparison overtime of ethnic group achievement in the Israeli elementary school. Evaluation in education, 1977, 1(1).

Lewy,A. and Crawford,W. Scoring a test battery. Educational and Psychological Measurement, 1968, 26(1), 70-77.

Lewy,A. and Kugelmass,S. (eds.) Decision oriented evaluation in education: the case of Israel. Philadelphia: International Science Services, 1981.

Lewy,A. and Nevo,D. (eds.) Evaluation roles in education. New York: Gordon and Breach, 1981.

Lewy,A., Rappaport,C. and Rimor M. Educational achievements in the Israeli school system. Tel-Aviv: Ramoth, 1978.

Lewy,L., Raz,R., Bachar,A. Summative evaluation of a science teaching curriculum for grades 7-8. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1983.

Lewy,A. and Raz,R. Achievement in Fourth Grade Mathematics. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1985.

Lockard,J.D. (ed.) Eight reports of the International Clearinghouse on Science on Mathematics Curriculum Developments. Maryland: College Park, University of Maryland, Science Teacher Center, 1972.

Luckey,J.W., Broughtois,A. and Sorensen,J.E. Archival data in program evaluation and policy analysis. Evaluation and Program Analysis, 1983, 5, 319-26.

Mager,R.F. Preparing instructional objectives. Palo Alto, Calif.: Fearon, 1962.

cClung,M.S. Competency testing: potential for discrimination. Clearinghouse Review, September 1977, 439-448 (ERIC Document Reproduction Service No. ED 164 643).

cClung,M.S. Are competency testing programs fair? legal? Phi Delta Kappan, 1978, 59, 377-400.

essick,S. Developing abilities and knowledge: style in the interplay of structure and process. Princeton, NJ.: Educational Testing Service, January 1983.(a)

essick,S. Assessment in context: appraising student performance in relation to instructional quality. Princeton, NJ.: Educational Testing Service, June 1983.(b)

inistry of Education and Culture. Guidelines for schools. Jerusalem: Ministry of Education and Culture, 1959.

inistry of Education and Culture. Mathematics syllabus for the kindergarden and for the elementary school. Jerusalem: Ministry of Education and Culture, 1972.

inistry of Education and Culture and Center of Educational Technology. One Two and ... Three: Textbooks (no. 1,2,3)- grade one. Tel-Aviv: Center of Educational Technology, 1979 (a).

inistry of Education and Culture and Center of Educational Technology. One Two and ... Three: Teacher's Guides (no. 1,2,3) - grade one. Tel-Aviv: Center of Educational Technology, 1979 (b).

inistry of Education and Culture and Center of Educational Technology. One Two and ... Three: Textbooks (no. 4,5,6)- grade two. Tel-Aviv: Center of Educational Technology, 1980 (a).

inistry of Education and Culture and Center of Educational Technology. One Two and ... Three: Teacher's Guides (no. 4,5,6) - grade two. Tel-Aviv: Center of Educational Technology, 1980 (b).

inistry of Education and Culture and Center of Educational Technology. One Two and ... Three: Textbooks (no. 7,8,9) - grade three. Tel-Aviv: Center of Educational Technology, 1981 (a).

Ministry of Education and Culture and Center of Educational Technology. One Two and ... Three: Teacher's Guides (no. 7,8,9) - grade three. Tel-Aviv: Center of Educational Technology, 1981 (b).

Ministry of Education and Culture and Center of Educational Technology. One Two and ... Three: Textbooks (no. 10,11, 12) - grade four. Tel-Aviv: Center of Educational Technology, 1982 (a).

Ministry of Education and Culture and Center of Educational Technology. One Two and ... Three: Teacher's Guides (no. 10,11,12) - grade four. Tel-Aviv: Center of Educational Technology, 1982 (b).

Minkovits,A. and Bashi,J. A proposal to reform the organization of supervision and guidance of teachers. Jerusalem: The Hebrew University, School of Education, 1973.

Minkovits,A., Davis,D., and Bashi,J. An evaluation study of Israeli elementary schools. Jerusalem: The Hebrew University, School of Education, 1977.

Moos,J.D. Towards equality: Progress by girls in mathematics in Australian secondary schools. Hawthorn, Victoria: Australian Council for Educational Research, 1982.

Movshovitz-Hadar,N. Simple Math. In: R. Hershkowitz (ed.) Proceedings of the Seventh International Conference for the Psychology of Mathematics Education. Rehovot: Weizmann Institute of Science, 1983.

National Assessment of Educational Progress. Math resource items for minimal competency testing: a collection of math items for state and local education agencies to drawn upon in custom-building their own minimal competency instruments. Denver: Education Commission of the States, 1977 (ERIC Document Reproduction Service No. ED 173 395).

National Council of Teachers of Mathematics. An agenda for action: Recommendations for school mathematics of the 1980s. Reston, Virginia: National Council of Teachers of Mathematics, 1980.

Nesher,P. Levels of description in the analysis of addition and

subtraction. Paper prepared for the seminar: The Initial Learning of Addition and Subtraction, University of Wisconsin, November 1979.

Nesher,P. The stereotyped nature of school word problems. For The Learning of Mathematics, 1980, 1(1), 41-48.

Nevo,D. The conceptualization of educational evaluation: An analytical review of literature. Review of Educational Research, 1983, 53, 117-128.

Nowakowski,J.R. On educational evaluation: a conversation with Ralph Tyler. Educational Leadership, May 1983, 24-29.

Nunnally,J.C. Psychometric theory. New York: McGraw-Hill, 1967.

Orthar,G. Educational achievements as related to socio-cultural background of primary school graduates in Israel. Megamoth, 1967, 15, 220-230. (Hebrew with English summary)

Padua,M. A survey on Bible Studies achievement. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1983.

Padua,M. Achievements in Biblical Studies in grade six and nine. Jerusalem: Ministry of Educationa and Culture, Israel Curriculum Center, 1986

Parish,R. and Arends,R. Why innovative programs are "discontinued". Educational Leadership, January 1983, 62-65.

Parlett,M. and Hamilton,D. Evaluation and illumination: a new approach to the study of innovatory programmes. Occasional Paper No. 9, Center for Research in the Educational Sciences, University of Edinburgh, 1972.

Patton,M.Q. Qualitative evaluative methods. Beverly Hills, Calif.: Sage, 1980.

Patton,M.Q. Evaluation for utilization. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Peaker,G.F. Sampling. In: T. Husen (ed) International study of achievement in mathematics: A comparison of twelve countries. Stockholm: Almqvist and Wiksell, 1967.

Peaker,G.F. An empirical study of education in twenty-one countries: a technical report. Stockholm: Almqvist and Wiksell, 1976.

Pedhazur,E.J. Multiple regression in behavioral research. (2nd edition) USA: Holt, Rinehart and Winston, 1982.

Pellerey,M. Mathematics instruction. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Pennsylvania Department of Educational. Educational quality assessment in Pennsylvania: the first six years. Philadelphia: 1973

Pophan,W.J. As always provocative. Journal of Educational Measurement, 1978, 15(4), 297-300.

Postlethwaite,T.N. Determination of general educational aims and specification of major objectives. In: A. Lewy (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.

Postlethwaite,T.N. International Association for the Evaluation of Educational Achievement. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Pring,R. Monitoring Performance: reflections on the Assessment Performance Unit. In: C. Lacey and D. Lawton (eds.) Issues in evaluation and accountability. London: Mcthuen, 1981.

Provus,M.M. Discrepancy evaluation. Berkeley, Calif.: McCutchan, 1971.

Raizen,S.A. and Rossi,P.H. Summary of program evaluation in education: when? how? to what ends? In: E.R. House, et al (eds.) Evaluation studies review annual (Vol. 7). Beverly Hills, Calif.: Sage, 1982, 31-52.

:az,R. A survey on Geography achievements. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1986.

:azel,C., Weisberger,M., and Sheinman,S. Identifying low achievers in reading and arithmetics in grade three. Jerusalem: The Henrietta Szoald Institute, 1984 (in Hebrew).

:esnick,D.B. Minimum competency testing historically considered. In: D.C. Berliner (ed.) Review of Research in Education (Vol. 8). Washington,D.C.: American Educational Research Association, 1980, 3-29.

:esnick,L.B. and Ford,W.W. The psychology of mathematics for instruction. Hillsdale,N.J.: Lawrence Erlbaum Ass., 1981.

.esnick,D.P. and Resnick,L.B. The nature of literacy: an historical exploration. Harvard Educational Review, 1980, 47, 370-85.

.iecken,H.W. Memorandum on program evaluation. (Internal memorandum to the Ford Fundation, 1953) In: C.H. Weiss (ed.) Evaluating action programs: readings in social action and education. Boston: Allyn and Bacon, 1972.

.iehs,R.J. (ed.) An international review of minimal competency programs in mathematics. Prepared for the International Study Group on Minimal Competencies in Mathematics, in cooperation with the Second International Mathematics Study - International Association for the Evaluation of Education Achievement (IEA). Columbus, Ohio: The Eric Science, Mathematics and Environmental Clearinghouse, Center for Science and Mathematics Education (The Ohio State University): March 1981.

iehs,R.J. Daily living skills. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

ossi,P.H. Boobytraps and pitfalls in the evaluation of social action programs. In: C.H. Weiss (ed.) Evaluating action programs: readings in social action and education. Boston: Allyn and Bacon, 1972.

ossi,P.H. Some dissenting comments on Stake's review. In: E.R. House, et al (eds.) Evaluation studies review annual (Vol.

7). Beverly Hills, Calif.: Sage, 1982, 224-235.

Rowley,G. Large-scale assessment and school-based curriculum: problems of compatibility. Paper presented to the 10th Annual Conference of the International Association for Educational Assessment. Perth, Western Australia, 11-15 June, 1984.

Sabar,N. Is the disappointment in the implementation of a new science curriculum justified? Journal of Research in Science Teaching, 1986, 23, 475-491.

Sarason,S.B. The culture of the school and the problem of change. Boston: Allyn and Bacon, 1971.

Scheier,M.A. and Rezmovitz,E.L. Measuring degree of program implementation: A methodological review. Evaluation Review, 1983, 7, 599-633.

Schildkampf-Kundiger,E. (ed.) Mathematics and Gender. Columbus, Ohio: The Eric Science, Mathematics and Environmental Clearinghouse, Center for Science and Mathematics Education (The Ohio State University): March- April 1982.

Scriven,M. The methodology of evaluation. In: R.E. Stake (ed.) Perspectives on curriculum evaluation. AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

Scriven, M. Maximizing the power of causal investigations: the modus operandi method. In: W.J. Pophan (ed.) Evaluation in education. Berkeley, Calif.: McCuthan, 1974.(a)

Scriven,M. Prose and cons about goal-free evaluation. Evaluation Comment, 1974, 3, 1-4.(b)

Scriven,M. How to anchor standards. Journal of Educational Measurement, 1978, 15(4), 273-275.

Sebring,P.A. and Boruch,R.F. How is national assessment: results of an exploratory study. Educational Measurements: Issues and Practice, Spring 1983, 16-20.

Shepard,L. Purposes of assessment. Studies in Educational Evaluation, 1979, 5, 13-26.

Shepard,L. Technical issues in minimum competency testing. Review of Research in Education, 1980, 8, 30-82.

Sherman,J.A. Sex-related cognitive differences: An essay on theory and evidence. Illinois: Thomas, Springfield, 1977.

Silberstein,M. Planning and empirical validation of a design for evaluating the implementation of a curriculum. Ph.D. Dissertation. Tel-Aviv: Tel-Aviv University, 1980 (in Hebrew).

Silberstein,M. Curriculum development in Israel in the Eighties: A dialogue between teachers and curriculum developers. Paper presented at the School Based Curriculum Development Conference. Tel-Aviv: Tel-Aviv University, January 1985.

Smith,E.R. and Tyler,R.W. Appraising and evaluating student progress. New York: Harper, 1942.

Soto,M.L. Teachers, parents, and community as data sources. In: A. Lewy (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.

St. Pierre,R.G. and Cook,T.D. Sampling strategy in the design of program evaluation. In: R.F. Conner, D.G. Altman, and C. Jackson (eds.) Evaluation studies review annual (Vol. 9). Beverly Hills, Calif.: Sage, 1982.

Stake,R.E. The countenance of educational evaluation. Teachers College Record, April 1967, 68(7), 523-540.

Stake,R.E. Language, rationality and assessment. In: M.W.H. Beatty (ed.) Improving educational assessment and an inventory of measures of affective behavior. Washington, D.C.: Association for Supervision and Curriculum Development, 1969.

Stake,R.E. Responsive Evaluation. Urbana: University of Illinois, 1972 (mimeograph).

Stake,R.E. A peer response. In: E.R. House, et al (eds.) Evaluation studies review annual (Vol. 7). Beverly Hills, Calif.: Sage, 1982, 53-60.

Stern,G., Stein,M.I.,Bloom,B.S. Methods in Personality Assessment. New York: Free Press, 1956.

Stuffleabeam,D.L. Evaluation as enlightment for decision making. In: M.W.H. Beatty (ed.) Improving educational assessment and an inventory of measures of affective behavior. Washington, D.C.: Association for Supervision and Curriculum Development, 1969.

Stufflebeam,D.L., Foley,W.J., Gephart,W.J., Guba,E.G.,Hammon,R.L., Merriman,H.O., and Provus,M.M. Educational evaluation and decision-making. Itasca,Ill.: Peacock, 1971.

Stufflebeam,D.L. Meta-evaluation. Occasional Paper No. 3. Kalamazoo: Western Michigan University, December 1974.

Suarez,T.M. Needs Assessment. In: A. Lewy and D. Nevo (eds.) Evaluation roles in education. New York: Gordon and Breach, 1981.

Suydam,M.N. and Dessart,J.D. Skill Learning. In: R.J. Shumway (ed.) Research in mathematics education. USA: National Council of Teachers of Mathematics, 1980.

Tamir,P. and Amir,R. Retrospective curriculum evaluation. Curriculum Inquiry, 1981, 11, 259-78.

Taylor,J. The foundations of maths in primary school. London: Allen and Unwin, 1976.

Templin,P. Photography in evaluation. Research on Evaluation Program, 1979, 2(1), 2-3.

The Joint Committee on Standards for Educational Evaluation. Standards for evaluation of educational programs, projects and materials. New York: McGraw-Hill, 1981.

The Henrietta Szold Institute, National Institute for Research in Behavioral Sciences. Report of the Committee for Examining the Utility and Feasibility of Establisihing a Feedback System. Jerusalem: The H. Szold Institute, 1981.

Thorndike,R.L. and Hagen,E. Measurement and evaluation in psychology and education. (4th edition.) New York: Wiley, 1977.

Tittle,C.K. Competency Testing. In: Encyclopedia of educational research (Vol 1 - 5th edition.), American Educational Research Association, 1982.

Travers,K. Mathematics: Key concepts. In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Travers,K. A summary of the IEA's second mathematics study. (in preparation).

Tobias,S. and Donady,B. Counseling the mathe anxious. Journal of National Association of Women Deans, Administrators and Counsellors, 1977, 41, 13-16.

Tyler,L.E. The psychology of human differences (2nd edition) New York: Appleton-Century-Crofts, 1956.

Tyler,L.E. Sex differences. In: R.L. Ebel (ed) Encyclopedia of Educational Research (4th edition) Toronto: MacMillan, 1969.

Tyler,R.W. Basic principles of curriculum and instruction. Chicago: University of Chicago, 1949.

Tyler,R.W. Some persistent questions on the defining of objectives. In: C.M. Lindvall (ed) Defining educational objectives. Pittsburgh, Pennsylvania: University of Pittsburgh Press, 1964.

Tyler,R.W. The purposes of assessment. In: M.W.H. Beatty (ed.) Improving educational assessment and an inventory of measures of affective behavior. Washington,D.C.: Association for Supervision and Curriculum Development, 1969.(a)

Tyler,R.W. (ed.) Educational evaluation: new roles, new means. Chicago: University of Chicago, 1969.(b)

Tyler,R.W. National Assessment of Educational Progress (NAEP). In: T. Husen and T. N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Van Bruggen,J.C. Establishing a modern core curriculum: a tricky business or a potential art? Paper presented at the

international seminar "Core Curriculum" in Western Societies. Enschede, Netherland, 12-15 Nov., 1985.

Vinner,S. New mathematics. In:T.Husen and T.N. Postlethwaite (eds.) The international encyclopedia of education: research and studies. Oxford: Pergamon Press, 1985.

Walker,D.A. The IEA six subject survey: an empirical study of education in twenty-one countries. Stockholm: Almqvist and Wiksell, 1976.

Walker,D.A. and Schafferzik,T. Comparing Curricula. Review of Educational Research, 1974, 44, 83-111.

Weiss,C.H. Evaluating educational and social action programs: a "treeful of owls". In: C.H. Weiss (ed.) Evaluating action programs: readings in social action and education. Boston: Allyn and Bacon, 1972.(a)

Weiss,C.H. Utilization of evaluation: towards comparative study. In: C.H. Weiss (ed.) Evaluating action programs: readings in social action and education. Boston: Allyn and Bacon, 1972.(b)

Weiss,C.H. The politicization of evaluation research. In: C.H. Weiss (ed.) Evaluating action programs: readings in social action and education. Boston: Allyn and Bacon, 1972.(c)

Weiss,C.H. Measuring the use of evaluation. In: E.R. House, et al (eds.) Evaluation studies review annual (Vol. 7). Beverly Hills, Calif.: Sage, 1982, 129-145.

Weiss,R.S. and Rein,M. The evaluation of broad-aim programs: difficulties in experimental design and an alternative. In: C.H. Weiss (ed.) Evaluating action programs: readings in social action and education. Boston: Allyn and Bacon, 1972.

Wiley,D.E. and Harnischfeger,A. Explosion of a mith: quantity of schooling and exposure to instruction, major educational vehicles. Educational Researcher, 1974, 4, 7-12.

Wilson,J.W. Secondary school mathematics. In: B.S. Bloom et al (eds) Handbook on formative and summative evaluation of student learning. New Jersey: McGraw-Hill, 1971.

Wolf,R. The identification and measurement of home environmental process variables that are related to intelligence. Ph.D. Dissertation. Chicago: University of Chicago, 1964.

Wolff, Y. The implementation of an innovative History curriculum. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1983.

Wolff,Y. and Padua,M. Centrally produced minimum requirement specification in subject areas and their impact on teachers behavior and learners' achievement. Jerusalem: Ministry of Education and Culture, Israel Curriculum Center, 1985.

Womer,F.B. and Martin,W.H. The national assessment of educational progress. Studies in Educational Evaluation, 1979, 5, 27-38.

Worthen,B. and Sanders,J.R. Educational evaluation: theory and practice. Washington,D.C.: C.A. Jones, 1973.

Yoloye,E.A. Observational Techniques. In: A. Lewy (ed.) Handbook of curriculum evaluation. New York: International Institute for Educational Planning, UNESCO, 1977.

Zak,I. Latent variables in causal models: an example from educational research. In: A. Lewy and S. Kugellmass (eds.) Decision oriented evaluation in education: the case of Israel. Philadelphia: International Science Services, 1981.

Zak,I. School's organization climate. In: A. Lewy and D. Nevo (eds.) Evaluation roles in education. New York: Gordon and Beach, 1981.

Zemer,H. and Raz,R. The summary of end of year tests for the "One, Two and ... Three" Program. Yedion, 1984, 47, 33-37 (Jerusalem: Ministry of Education, Curriculum Center, in Hebrew).

Zetterberg,H. On theory and verification in sociology. Tottowa, N.J.: Bedminster, 1965.

Zuzovski,R. The elementary school in Israel and science achievement. Thesis submitted for the degree Doctor of Philosophy. Jerusalem: Hebrew University, 1987 (in Hebrew).

Appendices

I	Distribution of the test items across contents and behavior dimensions.
II	Reliability coefficents of scales.
III	The data collection instruments (in Hebrew).

Appendix I

Distribution of the test items across contents and behavior dimensions*.

The behavior dimensions	The fourth grade mathematics contents				Total number of items
	Number System	Integers	Fractions	Geometry	
Computation		3, 4, 5, 11, 12, 13, 15, 36, 37, 38, 39, 40	24, 25	30	15
Comprehension	1, 2, 6, 28, 32, 33	7, 8, 14, 19, 26, 29	22, 23	34	15
Application		9, 10, 16 17, 27		18, 35	7
Analysis		20, 21, 31			3
Total number of items	6	26	4	4	40

*The figures appearing in the cells are serial numbers of the items, except the numbers in the marginal column and line which contain the "total number of items".

Appendix II

Reliability coefficients of scales

Name of scale	Code	No. of items	Reliability (coef. alpha)
*Professional background	BIO	6	.520
Autonomy	AUT	3	.527
Attitude to program	ATT	4	.559
Implementation	IMP	8	.477

*collapsed with "in-service training activities".

Appendix III

שאלון למנהל

לכבוד
מנהל בית־הספר

_____

אנו עורכים סקר על מצב הוראת המתמטיקה בבתי־הספר היסודיים, ובמיוחד בכיתות ד'. נודה לך אפוא, אם תקדיש כמה רגעים מזמנך למתן תשובות לשאלון קצר זה.

כתובתנו: האגף לתכניות לימודים, רח' פארן 7 ירושלים. טלפון 817251.

בתודה על שיתוף הפעולה,

צוות מתמטיקה — מדור הערכה
מלכה מאונטויטן — רות רז

כמה כתות ג' בבית־ספרך?_____

כמה כתות ד' בבית־ספרך?_____

לפי אילו ספרים מלמדים מתמטיקה בכל אחת מדרגות הכיתה בבית־ספרך בשנת תשמ"ז
(נא לסמן x במקום המתאים.)

סוג הספר	כיתות					
	א	ב	ג	ד	ה	ו
ספרי הרשקו						
ספרי תפר עד הדר						
ספרי "אחת שתיים שלוש"						
ספר "ועוד אחת"						
ספרי הטלוויזיה הלימודית						
אורבך "דרך הגילוי"						
אחר (ציין) _____						

תאם יש בבית־ספרך מרכז לצקצוע המתמטיקה? 1. כן 2. לא
(חקף בעיגול את התכון).
אם יש מרכז, רשום נא את שמו: _____

משרד החינוך והתרבות
האגף לחכניות לימודים
מקר מחימטיקה החשמ"ה

מבחן בחשבון לכיתה ד'
(סגור)

1. איזה מהמספרים הבאים הוא ארבעת אלפים שלוש מאות ושבע?

א. 4,037 ב. 40,307 ג. 4,370 ד. 4,307

---

2. מהו המספר הבא אחרי 599?

א. 590 ב. 600 ג. 598 ד. 601

---

3. חבר וסמן את התשובה הנכונה לתרגיל זה: 899 + 44

א. 483 ב. 1043 ג. 943 ד. 933

---

4. חבר וסמן את התשובה הנכונה לתרגיל זה: 2037 + 506

א. 543 ב. 2543 ג. 2533 ד. 2536

---

5. חסר וסמן את התשובה הנכונה לתרגיל זה: 5236 - 1824

א. 4412 ב. 6052 ג. 4312 ד. 3412

---

6. כפול וסמן את התשובה הנכונה לתרגיל זה: 40 × 200 =

א. 8,000 ב. 80,000 ג. 800 ד. 80

---

7. פתור וסמן את התשובה הנכונה לתרגיל זה: 3 × (6 + 2) =

א. 24 ב. 36 ג. 20 ד. 11

---

8. איזה מספר חסר במשבצת? □ × 5 + 30 × 5 = 31 × 5

א. 11 ב. 0 ג. 5 ד. 1

---

9. לכל שולחן ארבע רגליים. דני ספר עשרים רגליים. כמה שולחנות ראה דני?
סמן את התרגיל המתאים לשאלה זו:

א. 4 : 20 ב. 4 - 20 ג. 10 : 40 ד. 4 + 20

---

10. בקופת הכיתה היו 3,000 שקלים. לקראת מסיבת פורים הוציאו התלמידים 2,700 שקלים. כמה שקלים נשארו בקופה?
סמן את התרגיל המתאים לשאלה זו:

א. 2,700 : 3,000 ב. 2,700 - 3,000 ג. 2,700 + 3,000 ד. 2,700 × 3,000

---

11. חלק וסמן את התשובה הנכונה לתרגיל זה: 3 : 75 =

א. 15 ב. 215 ג. 25 ד. 27

---

12. חלק וסמן את התשובה הנכונה לתרגיל זה: 6 : 96 =

א. 26 ב. 34 ג. 36 ד. 16

13. כפול וסמן את התשובה הנכונה לתרגיל זה: $\begin{array}{r} 309 \\ \times\ 3 \\ \hline \end{array}$

א. 927 ב. 3027 ג. 117 ד. 327

---

14. איזה סימן יש לשים במשבצת הריקה $4 \square 100 = 25$

א. + ב. - ג. × ד. :

---

15. כפול וסמן את התשובה הנכונה לתרגיל זה: $\begin{array}{r} 76 \\ \times\ 7 \\ \hline \end{array}$

א. 492 ב. 502 ג. 532 ד. 512

---

16. כמה ספרים קנו חמישה ילדים בשבוע-הספר, אם כל ילד קנה שישה ספרים?
סמן את התשובה הנכונה לשאלה זו:

א. 36 ב. 11 ג. 22 ד. 30

---

17. רמי קנה 3 מדבקות שכל אחת 5 בולים, יוסי קנה 20 בולים. כמה בולים קנו יוסי ורמי? סמן את הביטוי הנותאים לשאלה זו:

א. $3 \times 5 + 20$ ב. $20 + 5 + 3$ ג. $5 + 20$ ד. $3 \times (5 + 20)$

---

18. נתונה יחידת המידה הזאת: ⊢——⊣
מה אורך הקו הזה? ⊢—+—+—+—+—+—+—+—+—⊣

א. 9 יחידות ב. 7 יחידות ג. יחידה אחת ד. 8 יחידות

19. אמא קנתה 1 ק"ג קמח. היא השתמשה ב-300 גרם לאפיית עוגה. כמה קמח נשאר?
סמן את התשובה הנכונה לשאלה זו:

א. 1 קילוגרם. ב. $\frac{1}{2}$ קילוגרם. ג. 700 גרם. ד. 600 גרם.

---

20. במגרש בית הספר שאול ספר 25 אופניים של בנות ופי שלושה מזה אופניים של בנים.
כמה אופניים לבנים?
סמן את התרגיל המתאים לשאלה זו:

א. 25×3 ב. $25 + 3$ ג. $25 + (25 \times 3)$ ד. 25×4

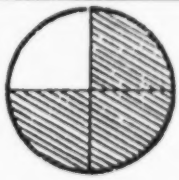
---

21. מחברת עולה 45 שקלים. אמא קנתה 3 מחברות לשרית ו-4 מחברות לדורית, כמה שלמה אמא? סמן את התשובה הנכונה:

א. 315 ב. 285 ג. 325 ד. 215

---

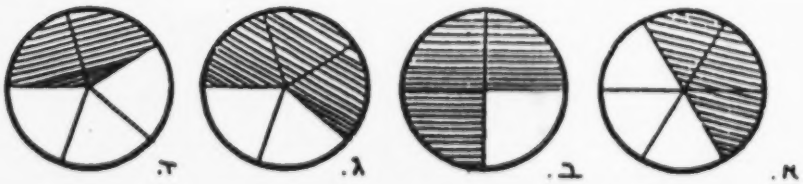
22. מהו השבר המתאים לשטח המקווקו?



א. $\frac{1}{4}$ ב. $\frac{3}{4}$ ג. 3 ד. $\frac{3}{8}$

---

23. באיזה עיגול צבעו $\frac{3}{5}$ נמנו?



א. ב. ג. ד.

24. חסר וסמן את התשובה הנכונה לחרגיל זה: $\frac{7}{9} - \frac{2}{9} =$

א. $\frac{5}{9}$ ב. $\frac{9}{9}$ ג. $\frac{4}{9}$ ד. $\frac{3}{9}$

---

25. חבר וסמן את התשובה הנכונה לחרגיל זה: $\frac{1}{2} + \frac{1}{4} =$

א. $\frac{1}{6}$ ב. $\frac{3}{4}$ ג. $\frac{2}{6}$ ד. $\frac{2}{4}$

---

26. השלם את המספר החסר כמשבצת: $\frac{\square}{6} = \frac{2}{3}$

א. 2 ב. 3 ג. 5 ד. 4

---

27. ליובל היו 17 גולות. במשחק עם דן הפסיד 9 גולות ובמשחק עם אסף הפסיד עוד 5 גולות, כמה גולות נשארו ליובל?
סמן את התשובה הנכונה:

א. 8 ב. 13 ג. 3 ד. 2

---

28. איזה מהמספרים הבאים הוא ארבע מאות ושמונה אלף, שלוש מאות עשרים ושתיים.

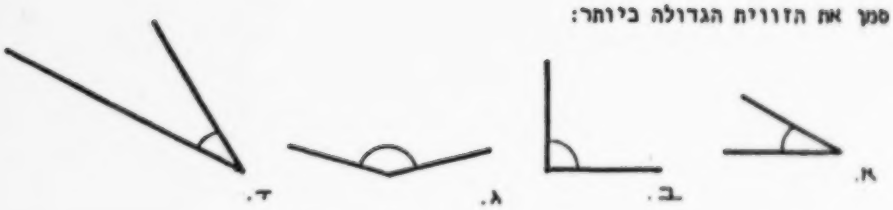
א. 480,322 ב. 408,322 ג. 48,322 ד. 4,008,322

---

29. סמן את המספר המתחלק ב-3 ללא שארית:

א. 155 ב. 143 ג. 171 ד. 229

30. סמן את הזווית הגדולה ביותר:



---

31. סמן בעיגול את האומדן הקרוב ביותר לפתרון של התרגיל 31 × 99 =

(אל תחשב את התשובה המדויקת):

א. 3,000 ב. 30 ג. 300 ד. 30,000

---

32. איזה מהמספרים האלה הוא מספר ראשוני?

א. 27 ב. 31 ג. 45 ד. 33

---

33. איזה מספר הוא פריק?

א. 3 ב. 13 ג. 23 ד. 33

---

34. באיזה מרובע כל הזוויות שוות?

א. מקבילית ב. מעוין ג. מלבן ד. דלתון

---

35. לגג הבית בנו גדר. אורכה של הגדר הוא 80 מטר ורוחבה 30 מטר. מה היקף הגדר מסביב לכל הגג?

א. 160 מטר ב. 100 מטר ג. 80 מטר ד. 60 מטר

36. כפול וסמן את התשובה הנכונה לתרגיל זה: $\begin{array}{r} 589 \\ \times\ 30 \end{array}$

א. 2010 ב. 17670 ג. 16770 ד. 619

---

37. כפול וסמן את התשובה הנכונה לתרגיל זה: $\begin{array}{r} 26 \\ \times\ 32 \end{array}$

א. 130 ב. 722 ג. 732 ד. 832

---

38. כפול וסמן את התשובה הנכונה לתרגיל זה: $\begin{array}{r} 245 \\ \times\ 106 \end{array}$

א. 3,920 ב. 25,970 ג. 3,510 ד. 351

---

39. חלק וסמן את התשובה הנכונה לתרגיל זה: 74 : 6 =

א. 12 ב. 12 ושארית 1 ג. 12 ושארית 2 ד. 12 ושארית 3

---

40. חלק וסמן את התשובה הנכונה לתרגיל זה: 44 : 5 =

א. 8 ושארית 4 ב. 8 ג. 9 ושארית 1 ד. 8 ושארית 5

משרד החינוך והתרבות
האגף לתוכניות לימודים

שם בית הספר _____
הישוב _____
שם המורה _____
(המלמד חשבון בכיתה ד')
הכיתה הנבחנת _____

שאלון למורה

סקר מתימטיקה החשמי"ת

(את תשובותיך עליך לסמן לרוב על-ידי הקפת עיגול סביב הספרה שליד התשובה המתאימה לך ולעתים על-ידי כתיבה מלאה של התשובה בגוף השאלון במקום המתאים)

א'-ד'. מהי השכלתך?

1. מורה מוסמך בוגר סמינר דו-שנתי.
2. מורה מוסמך בוגר סמינר חלת-שנתי.
3. בעל תואר אקדמאי מאוניברסיטה.
4. אחר: _____.

ה'-ח'. האם קיבלת הכשרה מיוחדת להוראת מתמטיקה?

1. לא.
2. כן, במוסד להכשרת מורים.
3. כן, באוניברסיטה.
4. כן, בקורס מרכזי מקצוע.

ט'. האם אתה מחנך כיתה ד' הנבחנת?

1. לא. 2. כן.

י'. האם אתה מרכז המקצוע מתמטיקה?

1. לא. 2. כן.

יא-יב. כמה שנים אתה מלמד מתמטיקה? _____.

יג-יד. כמה שעות שבועיות אתה מלמד בבית-ספר זה? _____.

טו. כמה שיעורים בשבוע מוקדשים להוראת מתמטיקה בכיתה ד' הנבחנת?

1. שלושה שיעורים או פחות.
2. ארבעה שיעורים.
3. חמישה שיעורים.

ספר הלימוד

טז-יז. לפי איזה ספר לימוד אתה מלמד השנה בכיתה ד' הנבחנת?_____

יח-יט. כמה שנים אתה מלמד לפי ספר זה?_____.

כ-כו. ציין לפי איזה ספר לימוד אתה מלמד או לימדת בדרגות כיתה א'-ג'.

1. לא לימדתי בדרגות כיתות אלה.
2. לימדתי בכיתה ג' (שם הספר)_____.
3. לימדתי בכיתה ב' (שם הספר)_____.
4. לימדתי בכיתה א' (שם הספר)_____.

כז-לב. ציין על פי אילו ספרי לימוד למדו תלמידי כיתה ד' הנבחנת השנים הקודמות.

דרגת הכיתה	שם הספר
כיתה ג'	
כיתה ב'	
כיתה א'	

לג-לז. מי מחליט באילו ספרי מתמטיקה ילמדו רבית ספרך?

1. המנהל.
2. המורה המלמד.
3. מרכז המקצוע.
4. המפקח.
5. אחר (פרט):_____

לח-מב. מי צריך לדעתך להחליט באילו ספרי מתמטיקה ילמדו בבית ספרך?

1. המנהל.
2. המורה המלמד.
3. מרכז המקצוע.
4. המפקח.
5. אחר (פרט):_____

מג. אם אתה עצמך בחרת בספר הלימוד שלפיו אתה מלמד השנה, סמן את הסיבה לבחירתך.

1. השתתפתי בהשתלמות והחומר נראה לי מתאים ומצא חן בעיניי.
2. השתתפתי בהשתלמות ולמרות הספקות החלטתי לנסות.
3. לא הייתי שבע רצון מהספר הקודם שלפיו לימדתי (ציין לפי איזה ספר לימדת)_____.

מד. האם השתתפת במשך השנה בישיבות מורים מיוחדות שבהן דנו על הוראת המתמטיקה בבית הספר?

1. אף פעם לא.
2. פעם אחת בשנה.
3. יותר מפעם אחת.

מה-מז. האם קיבלת הכשרה להוראה על פי גישה מסוימת במתמטיקה?

1. לא.
2. כן (פרט איזו גישה): _____.

מח-מט. באילו שנים: _____.

נ-נא. ציין כמה פגישות התקיימו: _____.

נב. האם השתתפת בהשתלמויות שהוקדשו להוראת ספר הלימוד שלפיו אתה מלמד השנה?

1. לא.
2. כן.

נג. כיצד התקיימו ההשתלמויות?

1. לא התקיימו השתלמויות.
2. רובן היו מרוכזות בחופשות ובימי עיון.
3. רובן היו מפוזרות במשך שנת הלימודים.

נד-נח. ציין באיזו מידה חשוב לעסוק בהשתלמויות בנושאים האלה (הקף בעיגול את הספרה בטור המתאים לדעתך, סמן לגבי כל נושא):

נ ו ש א י ם	1. כלל לא חשוב	2. חשוב במקצת	3. חשוב	4. חשוב מאוד
הסברת שיטות הוראה ודרכי הוראה	1	2	3	4
העשרת הנושא המתימטי והרחבתו	1	2	3	4
גישה לתלמידים ברמות שונות	1	2	3	4
ארגון העבודה בכיתה (כגון עבודה בקבוצות או עבודה אינדיווידואלית)	1	2	3	4
דרכי אבחון ובחינת	1	2	3	4

נט. האם אתה מקבל באופן שוטף הדרכה להוראת החשבון?

1. לא.
2. כן.

ס. אם כן, מי מבצע את ההדרכה?

1. אחד מהמורים בבית הספר.
2. מרצה (נציג פרוייקט) מן החוץ הבא לבית הספר.
3. ההדרכה ניתנת לקבוצת מורים מבתי ספר שונים במקום שמחוץ לבית הספר.

סא-סה. אם השתתפת בהשתלמות ציין באיזו מידה זכו הנושאים הבאים לטיפול משביע רצון. (הקף בעיגול את הספרה בטור המתאים לדעתך, סמן לגבי כל נושא)

ה נ ו ש א י ם	1. בכלל לא טופלו	2. טופלו במידה מועטה	3. טופלו במידה רבה	4. טופלו במידה רבה מאד
הסברת שיטת ההוראה ודרכי ההוראה	1	2	3	4
העשרת הנושא והרחבתו	1	2	3	4
גישה לתלמידים ברמות שונות	1	2	3	4
ארגון העבודה: בכיתה (כגון עבודה בקבוצות או עבודה אינדיווידואלית)	1	2	3	4
דרכי אבחון ובחינה	1	2	3	4

<u>מדריך למורה</u>

סו. האם אתה משתמש ב"מדריך למורה" במסגרת עבודתך בכיתה הנבחנת?

1. אין "מדריך למורה" לספר שאני מלמד לפיו.
2. יש "מדריך למורה", אך אני לא משתמש בו.
3. לא ידעתי על קיומו של "מדריך למורה".
4. ניסיתי לקנות "מדריך למורה" בחנות ולא הצלחתי.
5. ידעתי שלא אזדקק ל"מדריך למורה" ולכן לא ניסיתי להשיגו.
6. כן, אני משתמש ב"מדריך למורה".

(ענה על שאלה זו רק אם הנך משתמש ב"מדריך למורה").

סז. כיצד אתה משתמש ב"מדריך למורה"?

1. אני נעזר בו בהכנה כללית של הוראת הנושא.
2. אני נעזר בו בהכנת חלק מהשיעורים.
3. אני נעזר בו בהכנת כל שיעור.

<u>הערכת ספר הלימוד</u>

סח. באיזו מידה דרך הלימוד המוצעת בספר שלפיו אתה מלמד מתאימה לתלמידי כיתה ד' הנבחנת?

1. קשה לתלמידים.
2. קלה לתלמידים.
3. מתאימה לתלמידים.

סט. מה דעתך על הגרפיקה המופיעה בספר הלימוד?

1. מפריעה לתהליך הלימוד.
2. לא מוסיפה ולא מפריעה לתהליך הלימוד.
3. עוזרת לתהליך הלימוד.
4. אחר (פרט: _____)

ע. האם לדעתך ההוראות לתלמיד בספר הלימוד ברורות?

1. לא, בכלל לא ברורות.
2. לעתים אינן ברורות.
3. כן, ברורות מאוד.

עא. באיזו מידה רצוי לדעתך להקנות הרגלי עבודה בכלי הנדסה לתלמידי כיתה ד'?

1. בכלל לא.
2. במידה מועטה.
3. במידה רבה.
4. במידה רבה מאוד.

עב. איזה ציון מבטא באופן כללי את הערכתך לגבי ספר הלימוד שבו אתה משתמש בכיתה הנבחנת?

1. גרוע. 2. בינוני. 3. טוב. 4. טוב מאוד. 5. מצטיין.

מדוע?_____.

עג. הבעיה המילולית היא נושא חשוב בהוראת המתמטיקה. האם הטיפול בנושא זה בספר הלימוד משביע רצון?

1. אין טיפול בכלל.
2. יש טיפול, אך אינו משביע רצון.
3. יש טיפול משביע רצון.
4. אחר (אנא פרט):_____.

<u>חומר עזר (הרחבה)</u>

עד-עט. האם אתה נוהג להכין עזרים לימודיים נוסף לחומרי הלמידה הכלולים בספר?

1. לא.
2. כן, דפי עבודה.
3. כן, כרטיסי עבודה.
4. כן, משחקים.
5. כן, חומר להוראת הנדסה.
6. כן, אחר (פרט):_____

פ-פב. האם קיים שימוש בתכניות מחשב בהוראת מתמטיקה בכיתה ד' הנבחנת?

1. לא.
2. כן. ציין (איזו תכנית):_____

פג. אם התלמידים משתמשים בתכניות מחשב, האם יש תיאום בין תכנית המחשב לבין שיעורי המתמטיקה בכיתה?

1. לא.
2. כן.

פד. האם מותר לתלמידים בכיתה ד' הנבחנת להשתמש במחשבון (מחשב כיס)?

1. אף פעם לא.
2. כן, לפעמים.
3. כן, תמיד.

ארגון ההוראה

פז-פח. כמה שעות בשבוע צריך להקדיש תלמיד ממוצע בכיתה הנבחנת להכנת שיעורי בית?

_____.

פט. לפניך דרכים אחדות לארגון ההוראה בכיתה. סמן דרך אחת האופיינית ביותר למצב בכיתה הנבחנת.

1. כל התלמידים מבצעים פעולות זהות.
2. התלמידים לומדים בקבוצות הומוגניות.
3. התלמידים לומדים לפי רמות יכולת הטרוגניות.
4. התלמידים לומדים באופן אינדיווידואלי.
5. אחר (פרט): _____

צ. ציין איזה חלק מזמן ההוראה שלך בכיתה הנבחנת מוקדש לחזרה על נושאים שנלמדו בכיתות קודמות.

1. רבע מהזמן או פחות מזה.
2. בין רבע למחצית הזמן.
3. יותר ממחצית הזמן.

צא-צד. דרג את הפעילויות הבאות לפי סדר חשיבותן בעיניך (כתוב 1 ליד הפעולה החשובה ביותר 2 ליד הפעולה הפחות חשובה וכך הלאה).

זכירת עובדות מתמטיות _____

הוראה של טכניקות חישוב _____

פיתרון שאלות מילוליות _____

הבנת מושגים במתמטיקה _____

צה. האם מוכרת לך תכנית הלימודים במתמטיקה?

1. לא. 2. כן.

צו. באיזו מידה נעזרת בתכנית הלימודים כדי לקבוע את ספרי הלימוד ואת נושאי הלימוד?

1. בכלל לא נעזרתי.
2. נעזרתי במידה מועטה.
3. נעזרתי במידה רבה.
4. נעזרתי במידה רבה מאוד.

אם יש לך הצעות לגבי שינויים רצויים בתכנית הלימודים של תלמידי כיתה ד'.

פרט אותן: _____

_____

לפניך רשימת הנושאים כפי שהם מופיעים בתכנית הלימודים החדשה במתמטיקה לכיתה ד'.
אנא הקף בעיגול את הספרה בטור המתאים לדעתך, ביחס לגבי כל נושא.

1. אם התלמידים לא למדו אותו בכלל.
2. אם התלמידים למדו אותו בשנה שעברה בכיתה ג'.
3. אם התלמידים למדו אותו השנה בכיתה ד'.
4. אם התלמידים עדיין לא למדו אותו אך ילמדו אותו אחרי מועד המבחן, השנה או בשנה הבאה.

	נ ו ש א	1. לא למדו בכלל	2. למדו בכיתה ג'	3. למדו השנה	4. ילמדו אחרי מועד המבחן
צז.	הכרת מספרים גדולים (חזרה)	1	2	3	4
צח.	שיטת הא"ב (עד ח')	1	2	3	4
צט.	חיבור וחיסור "ללא הגבלה"	1	2	3	4
ק.	חכונות ה-0 וה-1	1	2	3	4
קא.	מספרים ראשוניים ופריקים	1	2	3	4
קב.	סימני התחלקות ב-3, 6, 9	1	2	3	4
קג.	אומדן	1	2	3	4
קד.	כפל בטור	1	2	3	4
קה.	חילוק ארוך	1	2	3	4
קו.	בעיות מילוליות	1	2	3	4
קז.	ציר המספרים, חיבור שליליים ויישומים הכרת השבר	1	2	3	4
קח.	שמות שונים לשבר, שבר מעורב, חיבור וחיסור, שברים עם מכנים שווים או קרובים וכפל שלם בשבר	1	2	3	4
קט.	מחשבון (תוך שימוש באומדן)	1	2	3	4
ר.	הסתברות/חזקות	1	2	3	4
רא.	חזרה	1	2	3	4
רב.	הכרת חיבות	1	2	3	4
רג.	נפח חיבות	1	2	3	4
רד.	זמן (ימים, שבועות, חודשים)	1	2	3	4
רה.	משפחת המרובעים (כולל חכונות הסימטריה)	1	2	3	4
רו.	טרטוט	1	2	3	4
רז.	סימטריה	1	2	3	4

תודה!

שאלון לתלמידים בכיתה ד'

(הקף בעיגול את המספר שליד התשובה שלך.)

כא. האם אתה בן או בת?
1. בן. 2. בת.

כב. האם אתה אוהב ללמוד עברית?
1. לא אוהב. 2. לא כל כך אוהב. 3. אוהב. 4. אוהב מאוד.

כג. האם אתה אוהב ללמוד חשבון?
1. לא אוהב. 2. לא כל כך אוהב. 3. אוהב. 4. אוהב מאוד.

כד. האם אתה אוהב ללמוד התעמלות?
1. לא אוהב. 2. לא כל כך אוהב. 3. אוהב. 4. אוהב מאוד.

כה. האם אתה אוהב ללמוד מולדת?
1. לא אוהב. 2. לא כל כך אוהב. 3. אוהב. 4. אוהב מאוד.

כו. האם אתה אוהב ללמוד תנ"ך?
1. לא אוהב. 2. לא כל כך אוהב. 3. אוהב. 4. אוהב מאוד.

כז. האם אתה אוהב ללמוד טלאכה?
1. לא אוהב. 2. לא כל כך אוהב. 3. אוהב. 4. אוהב מאוד.

כח. האם ספר הליבוד שלך בחשבון מעניין?
1. לא. משעמם מאוד. 2. לא. די משעמם. 3. מעניין. 4. מעניין מאוד.

כט. האם קל לך ללמוד בספר הליבוד בחשבון?
1. קל. 2. קצת קשה. 3. קשה. 4. קשה מאוד.

ל. האם ספר הליבוד שלך מלמד דברים חשובים?
1. לא, הם לא חשובים. 2. כן, הם חשובים. 3. כן, הם חשובים מאוד.

לא. האם כדאי להוסיף בכל שבוע שיעורים נוספים לליבוד חשבון?
1. לא. 2. כן.

לב. האם לימוד החשבון עוזר לנו בחייב?
1. לא. 2. כן.

לג. האם הליבודים בחשבון קשים לך יותר מהליבודים בטקצועות אחריב?
1. קשים בהרבה. 2. קשים בטקצת. 3. בכלל לא קשים.

לד. כשבתבטל שיעור חשבון, איך אתה מרגיש?
1. עצוב. 2. לפעמים עצוב ולפעמים שמח. 3. שמח.
מדוע?_____

לה. האם היית רוצה לעבוד בטקצוע שטשתמשים בו הרבה בחשבון?
1. לא. 2. כן. 3. אינני יודע.

לו. האם הוריך רגילים לשוחח אתך על ליבודי החשבון?
1. אף פעם. 2. לפעמים. 3. תמיד.

לז. האם בעיני הוריך חשוב מאוד שתצליח בליבוד החשבון?
1. לא. 2. כן.

לח. אם יש צורך לבטל שיעור,איזה שיעור הרצת שיבטלו? (בטן רק בקצוע אחר.)
1. חשבון. 2. עברית. 3. תנ"ך. 4. התעמלות. 5. טולדת 6. טלאכה.

לט. כשהטורה טודיעה שבטקום שיעור תנ"ך יש שיעור חשבון, איך אתה טרגיש?
1. עצוב. 2. לפעמים עצוב ולפעמים שמח. 3. שמח.
מדוע?_____

מ האם אתה מפחד משיעורי חשבון?
1. מפחד מאוד. 2. קצת מפחד. 3. בכלל לא מפחד.

מא. האם יש לך בחשבון (מחשב-כיס)?
1. לא, בכלל אין לי. 2. לא, אבל בכיתה נותנים לנו. 3. כן, יש לי.

מב. האם אתה משתמש בחשבון?
1. לא אינני משתמש. 2. כן, בבית. 3. כן, רק בכיתה. 4. כן, בבית ובכיתה.

הודה!

מסת"ב : 1—021—262—965